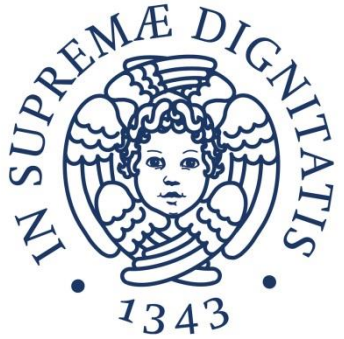


UNIVERSITA' DI PISA

Department of Economics and Management

Master of Science in Economics

Patents sharing knowledge with papers:
An empirical study on medical research when scholars turn into inventors

SUPERVISOR:
Giovanni Dosi
Arianna Martinelli

CANDIDATE:
Elena Romito

Academic Year 2014/2015

Acknowledgements

I would like to thank Arianna Martinelli for her guidance, patience, support and knowledge throughout this work. I am grateful to Giovanni Dosi for his supervision and advises. I also wish to thank Alessandro Nuvolari for his precious insights.

Abstract

This work aims to investigate how patent strategies in the field of medical innovation can influence the long-run production of public knowledge (i.e. ideas disclosed through open common institutions). In order to tackle this issue, we identify pieces of knowledge disclosed thorough both papers and patents and we estimate the effect of the granting on the number of citations to the paper. Since the patent grant comes with a lag of 3-4 years after the paper publication, it is possible to see the grant as an exogenous shock for the paper. We used the USPTO DATABASE for data on patents and, specifically, technological class 435/7.23 -which collects all patents referring to *tumor cells and cancer cells*- constitutes the original set. Data about papers, instead, are retrieved from the ISI WEB OF SCIENCE. Our results show that the effect of granting property rights on basic research for cancer is on the type of authors citing the publication (i.e. patents' inventors vs. third parties), rather that on the total amount of citations.

Contents

List of Figures

List of Tables

Introduction

Absolute certainty is not a feature characterising economics. What can be obvious and sure for one, can turn into a paradox for another. However, no schools of thought deny the central role of knowledge for progress and economic growth. Hence, a discussion about the possible ways for spurring innovative activities is highly relevant in economic theory.

Considering intellectual property rights (IPRs) as the first mechanism to stimulate research efforts has been an idea widely accepted and rarely challenged. However, during the last three decades, more critical outlooks and arguments have emerged. In particular, some authors[1] have suggested that when new knowledge is protected by a too strong monopoly (i.e. a very strict patent system), it may cause a decline of the follow-on rate of innovation. This may hamper the innovative process as this is very cumulative and the past innovative steps are typically the main ingredients for the production of new pieces of knowledge.

Although patents were born as a tool to remunerate efforts conducted by private institutions, nowadays these are widely used even in the public sector. The Bayh-Dole Act -signed in 1980- is conventionally considered the main policy change responsible for such a transition. In fact, it conferred to publicly funded Universities and other public institutions the right to cover with patents their discoveries. In this work we will deal with the concepts of *proprietary knowledge* as opposite to *public knowledge*. The first refers to information disclosed in exchange of property rights (patents), the second, instead, relates to ideas available in public platforms of disclosure (i.e. papers published in scientific journal). With changes in the legal environment and

[1] See for example Heller and Eisenberg, (1998); Bessen and Meurer (2008)

consequently shifts in patent practices, a clear distinction between subjects involved in one sphere instead of the other becomes blurred. The traditional linear model (Bush, 1945) that establishes a relation that goes from basic research (originated in public institutions) to more marketable innovations (due to the firms efforts), is not suitable to describe a modern innovation path.

A broad set of studies[2] have considered the impact that *public knowledge* has on *proprietary knowledge*[3] recognising the fundamental importance that the first has on the latter. On the contrary, a scarce attention has been given to the opposite relation. Among the few studies focusing on the effect that private tools of disclosure have on the rate of public knowledge, particularly relevant is the Huang's and Murray's (2009) work. In that paper they found out that the practice of patenting is strongly detrimental for future public knowledge in the field of genetics.

In this work, we are going to focus our attention on the field of research about *tumor and cancer cells*. The reason motivating this choice is twofold. Firstly, themes dealing with medical and therapeutic conditions, assume a relevance not only at the political and economic level, but these also bear ethical implications. Secondly, processes of medical research are among those ones that have mostly exploited the practice of covering with patents pieces of knowledge classifiable as basic information, rather than applicable and easily marketable.

The final aim of this work is finding an answer to the following research questions:

- In which way knowledge disclosed by both patents and publications differs from information disclosed only by means of private tools?
- Is the path of follow on knowledge influenced by the granting of a patent associated to the same piece of knowledge?

[2]For a further discussion on this matter see for example Cockburn and Henderson (1998); Cohen and Levinthal (1990)

[3]Note that in the literature the notion of *private knowledge* is often used as opposite to the concept of *public knowledge*. However, we prefer to define the patented information as *proprietary knowledge* since it is available for agents different from patent's inventors. On the contrary, *private knowledge* refers to secret and not disclosed information.

The basic assumption that must be satisfied in order to deal with this issues is that there exist some pieces of knowledge originally disclosed in scientific papers and, later, covered by a patent (*patent-paper-pair* form of disclosure[4]). In cases when it occurs, it is possible to isolate the effect that the patent grant has on the *following life* of the publication.

Work-plan

In order to achieve our final aim we are going to articulate our work thorough the following steps.

Chapter 1 presents a review of the main important theories that historically have aimed to outline the nature of technical progress. Besides the classical *Demand-pull* and *Technological-push theory* we in depth describe the features of an *evolutionary theory* where the procedural dimension of technical knowledge is considered to be a cornerstone. Therefore, we will argue that, when an evolutionary view is accepted, the set of mechanisms able to guarantee returns for innovative activities is actually much broader than the one considered in a classical framework.

Chapter 2 is going to investigate the features of the current U.S. patent system outlining both historical and theoretical foundations of IPR regimes. Later, we conduct an exhaustive review on the criticisms pointed out in the literature. Beyond the theoretical argumentations we consider the main empirical evidence either confirming or denying those criticisms.

Chapter 3 will be focused on our specific sector of interest: medical innovation. After a brief explanation about the evolution of the industrial structure, we will propose a review of the main features characterising the innovative process carried out by both private and public institutions. The first ones typically invest in R&D aiming at the production of those innovations closer to market. The second, instead, focus their attention on the upstream level of research. Therefore, at the end of the chapter, we rely on the existing literature for the sake of comprehension about the possible effect

[4]The concept of disclosure by means of *patent-paper-pairs* has been introduced in Murray (2002) and it has been further implemented in the author's following work (See Huang and Murray (2009))

of patents when they are granted on basic medical research.

Chapter 4 explains the methodology we used to conduct our empirical analysis and supplies the definitions of the entire set of variables adopted in the case of study. Specifically, we dedicate a section to an in depth explanation of *patent-paper-pair* approach giving a substantial importance to the identification strategy used to isolate pieces of knowledge simultaneously disclosed by patents and papers. Later we describe the statistical tools we applied for conducting both inferential and descriptive analysis.

Chapter 5 contains all the results we obtained at the end of our study. In opposition with our original expectations, at a first glance, we found no evidence that the practice of patenting has a significant effect on the amount of following public knowledge. However, the unexpected results gave us the right spur to deepen our analysis and it led to conclude that even though the effect on the rate of forward paper's citations (used as proxy for the rate of follow-on public knowledge) is not clear, the nature of agents using the piece of knowledge changes after the grant.

Finally, in the last part of our work we discuss the obtained results in terms of policy implications and we propose some possible directions for future research agenda.

Chapter 1

Innovation in evolutionary terms

The strict relation between economic growth and technical progress is evident and well recognised in economic thought[1]. However, the nature of this relation and the causality rules are issues much more controversial. This chapter will try to describe the debate on this matter justifying a preference for an evolutionary view of the innovative process.

1.1 Why does Innovation occur?

The first question that needs to be faced in order to describe the nature of the technical progress is about what triggers innovative process. On this matter, the two theories that have historically assumed a crucial importance are the *Demand-pull theory* and the *Technological-push theory*, as recognised in Dosi (1982).

Demand-pull theory

The basic idea besides this theoretical approach is that market forces are the first mover for inventive activity and technological progress. Rosenberg and Mowery (1979) describe in detail the various steps that, according to a marked-centred view, mark the path of the technological progress. The au-

[1] Dosi (1982), *Technological paradigms and technological trajectories*,p. 147

thors conduct this analysis in response to a series of empirical studies[2] that seem to undoubtedly confirm the centrality of markets force for spurring innovation. Their critical conclusion, however, is that the proposition according to which market demand governs the innovation process is by no means conclusively demonstrated in these studies. Before pointing out the possible causes behind this advocated invalidity, it is worth considering the basic points supporting the theory. At a given time, there exists a set of consumption and intermediate goods satisfying different needs. Consumers express their preferences about the features of the good they desire by means of their pattern of demand. When the budget constraint relaxes (because a growth in income), users' demand will grow more than proportionally for those goods having relatively preferred characteristics. At this point, producers recognize the revealed needs of the consumer and the innovative process begins as a response to such a need [3].

The criticisms that can be moved to such a strict and pure theory are quite evident. The first one emerges by the definition of need itself. It is difficult to accept, indeed, that agents are able to recognise their needs still before than a solution for that need actually exists[4]. Moreover, even when an *ex-ante recognition* is possible, needs are potentially infinite and it makes particularly hard to understand why a certain invention occurs in a specific point of time, instead of another.

Beyond the ambiguous concept of need, also the idea of technology appears to be too versatile and responsive in this context, like if it was a *freely available blackbox* (Dosi (1982), p. 150). In fact, even if it was possible to assert the necessity of creating something that the market wants, it would still be very difficult to invent it without considerable efforts and hight costs. Despite these evident criticisms, the *Demand-pull theory* has been advocated as the most reliable one for a long time and a quite supporting evidence has

[2]See for example Langrtsh J. (1972) and Meyers S. and Marquis D.G. (1969)

[3]This process has been well described in Dosi (1982), pag 149

[4]This statements is obviously not true for a set of primary needs that it would be possible to recognise even without knowing a about a solution for them (think, for example of eating, drinking or sleeping)

been proposed[5]. One possible reason behind this apparent contradiction was given by Rosenberg and Mowery (1979). In fact, they suggest that the theory has been tested only in a scenario of successful innovation. The inventive process, however is made of trials that can end in success, but, more often these just reveal to be failures. The latter dimension is rarely consider in *pure-market-centred-models*.

Technological-push theory

This second approach may be fitted, somehow, at the antipodes of the previous one. Here, the innovative process does not depend in any way from market requests, but, on the contrary, it would be consequent to some exogenous factors determining technological changes. Among them the increasing role of scientific inputs and the increased complexity of R&D activities are the most relevant ones [6]. The main consequence of considering progress in science as the only responsible for a shift in the technological path is that it becomes totally independent from the economic and social environment. Moreover, this view considers only the a downward relationship that accounts for the impact of *pure science* on *applicable technology*. However, these is a non scarce set of cases where the opposite relation occurred[7]. Therefore, although the importance of science for technological progress is undeniable, a linear model that goes from the top to the bottom of the chain cannot be established [8].

An alternative and evolutionary theory

The shortcomings of the previous theories leave the desire of finding an approach able to better understand the dynamic and complex nature of the innovative process. Evolutionary economics explain that the main failure of the traditional theories is the definition of technology itself. In a neoclassi-

[5]See for example Langrish (1972), Myers (1969)

[6]Dosi (1982), pag.151

[7]In XVII, the ship design case is a good example where the technological advantages were mostly due to the actual experience rather than to theoretical science

[8]For further discussion, see Bush 1945

cal framework technology may be defined as *a given set of factor combinations, defined (qualitatively and quantitatively) in relation to certain output.* It follows that the technological progress can be seen as *a moving production possibility curve and/or in terms of the increasing number of producible goods*[9].

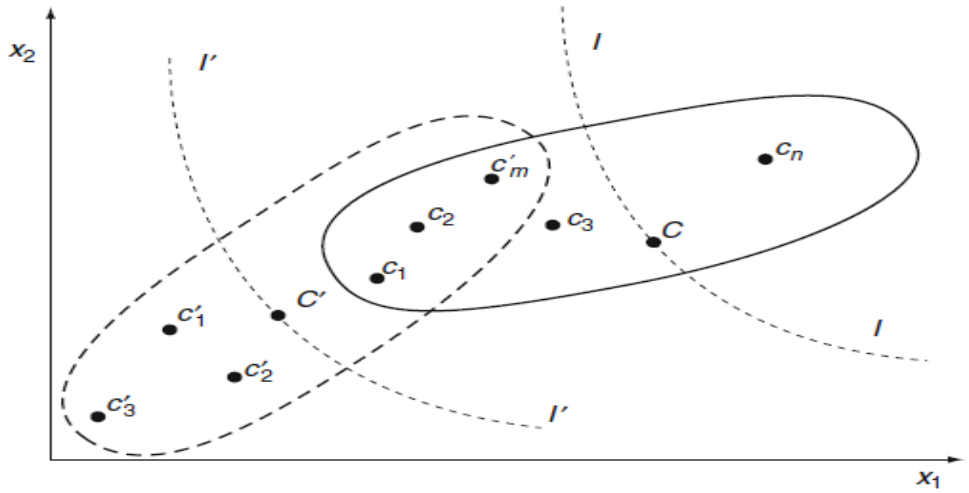
In an evolutionary view, these definitions are very far from being satisfying. The first definition of evolutionary economics dates back to 1898 when Veblen proposed the following interpretation of the economic theory:

> *A theory of a process of cultural growth as determined by the economic interest, a theory of a cumulative sequence of economic institutions stated in terms of the process itself.*(Veblen (1898), pag.389)

It is clear that in a such broad view of the economic theory, also the idea of technology has to be evaluated in wider terms. On this stream, Dosi (1982) defines technology as *a set of pieces of knowledge; both directly practical and theoretical.* Here the practical dimension relates to the development as a process of problem-solving activity. Whereas, the theoretical level is referred to the experience of the past attempts and to the knowledge of the state of arts. Starting from this view of technology, the author establishes a powerful parallelism between science and technical process. Khun (1962) introduced the concept of *scientific paradigm* defined as *an outlook which defines relevant problems, a model and a pattern of inquiry*[10]. Analogously, Dosi outlines the definition of *technological paradigm* as *a model and a pattern of solutions of selected technological problems, based on selected principles derived from natural science.* Moreover, an additional comparison has been made between the *normal science* and *technological trajectories.* As the normal science is *the actualization of a promise*, the technological trajectory has been described as *the pattern of 'normal' problem solving activity on the ground of a technological paradigm.* In other words, a technological paradigms identifies the generic tasks, the material technology and the available scientific proprieties (See 1.1). Once a paradigm has established,

[9]Dosi (1982), p. 151
[10]Khun (1962), pag. 23

Figure 1.1: Technological paradigms and trajectories



Source: Dosi and Nelson 2010

different possible technological trajectories can develop. At this point, in order to answer the question *Why does Innovation occur?* is not enough focusing on the various causes for shifting in the production function, but, on the contrary, the idea of a mere functional link between inputs and output is refused. The right question, in an evolutionary view, is *Why do some trajectories take the advantage on other possible ones?* In order to give an exhaustive answer neither reactions to markets requests, nor changes coming from outside are features that, taken separately, have the power to explain the development of technological process. At different levels, indeed, both endogenous and exogenous movers are fundamental for innovative activities. Keeping in mind the parallelism science-technology, a chain that goes from the pure science to market dimension can be imagined. The process of innovation, therefore, has to be analysed not only at the firm level (that is actually only the lowest one), but it clearly begins in a more abstract dimension. At this early stage the economic interests of organizations involved in R&D activities and the interest of public agencies are much more effective than market mechanisms which, therefore, loose their ex-ante power of selection. On this matter, the most explanatory example may be the computer case in the II post-war when government institutions were the primary source

of founds to be invested in R&D [11]. Moreover, the discussion on the role that public institutions play in the innovative activity leads to another important aspect that has historically been central in the debate about the position that single and *heroic entrepreneurs* perform in the technological process[12]. If at the primary stage of breakthrough innovation, the importance of public founding is hardly questionable, at a firm level Malerba and Orsenigo (2000) distinguishes *Schumpeter mark I* from *Schumpeter mark II* models. In the first case innovation results from the fragmented efforts of a myriad of small enterprises, the latter are characterised by highly concentrated R&D activities by few large and mostly public corporations. The prevalence of a model on the other strictly depends on the nature of the industrial sector [13]. The central features that determines the most applicable model is the degree of *cumulativeness* in different technological regimes. The property of *cumulativeness* captures the degrees to which *success breeds success*, that is the measure to which innovative advances are made by dwarfs standing on the shoulders of past giants and it varies a lot across different innovative activities[14]. A process is much more cumulative as the knowledge involved in it is endogenous. It is rather intuitive, at this point, that the higher the degree of *cumulativeness* is, the harder the learning path becomes for those outside the inventive sector. Therefore, in presence of an highly endogenous learning process, *Schumpeter mark II* models will apply. On the contrary, when learning is not easy, but at least possible, for entrants *Schumpeter mark I* models are more likely to occur.

A part from the above discussed distinction, evolutionary economists agree in admitting the importance of the economic and social environment in shaping the long-run direction of the technological process. In fact, although the idea of a black-box technology capable of instantaneously react to changes in the demand is strongly refused, users' value and their requests

[11] For further information on the role of government institutions on innovation, see Freeman and Soete (1997)

[12] The term heroic entrepreneur has been coined by Schumpeter (1934) and it refers to those entrants that, thanks to their special skills bring new innovation into the market, destructing the value of the old ones.

[13] See Dosi, Marsili, Orsenigo and Salvatore (1995)

[14] See Dosi and Nelson (2010), p.73

cause an important *inducement effect*[15]. The latter operates mostly through two channels. Firstly, there can be an impact on the orientation of search as described in Rosenberg (1976)[16]. Secondly, the effect can be on the intensity of search (rather than on its direction) in accordance with the *Schmookler's hypothesis* (Schmookle, 1966).

Summing up, from an evolutionary standpoint the perception of the mechanisms behind inventive activity becomes much more articulated and not easily explainable. As a consequence, it is impossible to re-conduct under a unique set of factors the principal movers for innovation. The complexity of the process is basically due to the nature of technology itself that is not seen any more as a black-box but as a set of pieces of knowledge. However, the latter is just a broad definition that can be better understood when other interpretations are considered.

1.2 Alternative interpretations for technology

Dosi and Nelson (2010) identify four different (but complementary) ways to describe the concept of technology. In this section[17], these will be briefly explained in order to underline the *procedural dimension* of technology, which becomes the main feature for the analysis of the appropriability conditions.

Technology as information

The parallelism between technological knowledge and information can be established on the bases of same shared characteristics: (i) non rivalry[18] in use (ii) indivisibility[19] (iii) notional scale-free property. The latter is referred to the property for which information *stricto sensu* has typically very low cost of

[15] See Dosi and Nelson (2010), p.75

[16] An example showing this effect in practice are the experience of the resistance of nineteenth-century English labour, to factory discipline. For further discussion on this matter see Rosenberg (1976)

[17] Dosi and Nelson (2010), pp.56-63

[18] The use of one economic agent of a piece of knowledge, does not reduce the possibility for another agent to use the same

[19] Note that in presence of non-divisible commodities, the basic neoclassical assumption of convex consumers' consumption set drops (Mas-Colell, 1995).

reproduction if compared with the hight generation costs. The implication is that above a certain threshold, information can be used at any level (with no depreciation) of production without incurring in the neo-classical limitation of constant return to scale.

However, it has to be stressed that property (iii) holds for technological knowledge only when it is considered in its *pure sense*, whereas, in reality, at least two characteristics weaken the notional scale-free hypothesis. Firstly, although it keep being true that the generation costs are much higher than the reproduction costs, these are not negligible[20]. The *cost of teaching* is just an example among them. Secondly, technological knowledge is characterized by several degrees of *tacitness*. The concept is well expressed by Pavitt:

> *Most technology is specific, complex. . .[and] cumulative in its development. . . It is specific to firms where most technological activity is carried out, and it is specific to products and processes, since most of the expenditures is not on research, but on development and production engineering, after which knowledge is also accumulated through experience in production and use on what has come to be known as 'learning by doing' and 'learning by using. [...] The combination of activities reflects the essentially pragmatic nature of most technological knowledge. Although a useful input, theory is rarely sufficiently robust to predict the performance of a technological artefact under operating conditions and with a high enough degree of certainty, to eliminate costly and time-consuming construction and testing of prototype and pilot plant* (Pavitt (1987), p.9).

Therefore, pre-existing knowledge necessary to apply any codified information, implies non trivial costs to acquire the necessary capabilities for using new technologies. The tacit components of new technological processes have such a strong relevance, that even in those (few) cases when an Arrow core[21] is present, the reproduction efforts are still significant. Moreover, the degree of uncertainty about the ultimate success does not vanishes when the

[20] For a detailed discussion on this matter see Manfield et al (1981)

[21] The Arrow core concept has been described in Winter and Szulanski (2002). Here the authors define Arrow core as an informationally codified template

know how component of the process takes the advantage on the *written and transferable* part.

Technology as recipe

An other effective way of describing technology is considering it as *a sequence of cognitive and physical acts that has to be performed in order to achieve a final product*[22], exactly as a recipe for a cake. Therefore, not only the ingredients (inputs) must be specified, but also the entire set of procedures that are necessary to get the final output. This procedure-based view of technical process in another important break point between evolutionary and traditional theories. The second ones (over)use the production function as a tool to quantitatively describe a process. However, the list of ingredients, alone, is never enough to describe an output. Moreover, when the procedural level is considered, one of the concept at the basis of the entire neo-classical framework is challenged: the is the substitution effect. Changing in the prices of inputs, indeed, cannot imply an immediate *rebalancing* of their quantities used in production, since variations in the ingredients, necessary require some modification in the procedural part of the recipe as well.

It is important to notice that, even considering technological knowledge as recipes, the tacit component of the process still play a crucial role. Obviously, the more a recipe is detailed, the more the final result will be close to the original one. However, your grandmother's cake will always be better than yours.

Technology as routine

Often one single person is not able to put in practice what a recipe requires. The various pieces of knowledge, indeed, are usually distributed across many individuals. Therefore, different people and group are assigned different parts of the process. This apparently trivial observation in actually introducing another important dimension of technological progress still never mentioned

[22]See Dosi and Nelson (2010), pag. 59

so far, i.e. the concept of *social technology*. It can be defined as *The system of norms, belief and social practices shaping the way of doing things*[23].

The standpoint of technologies as routines considers simultaneously the physical and the social levels of the process. Recipes alone, indeed, represent only the series of *legal procedures*, that is to say procedures that are technically feasible. The division of labour and specific models of coordination, instead, are part of a broader concept of technology which is well described by the idea of routines. It worth noticing, at this point, that the introduction of a new technique, not only has an impacts on inputs and on procedures, but it often requires a change also in *rutinezed path*, it ,makes the scenario even more complex.

Technology as atrifact

When the output of a recipe is an a physical complex good, the procedure-centred description of technology becomes broadly complementary to an artifact-centred view. In the second case, the final artifact can be considered as the technology itself. In particular, it is possible to identify the final good with the technical process behind, by means of two different approach. Firstly, the identification of the techno-economic characteristics of a specific final good. It allows for tracing an history of technology that, it has been argued, shows an *hedonic dimension* of the innovation[24].
Secondly, the different components of the good and the necessary intermediate inputs can be separately analysed in order to describe the technical path behind a specific artifact.

1.3 Means of appropriation

When considered together, the aforementioned notions of knowledge lead to refuse the idea that knowledge can be considered as a pure common good. In fact, although it keeps being non-rival in use, the predominance of the tacit,

[23]Nelson and Sampat (2001)
[24]Dosi and Nelson (2010), p. 62

procedural and rutinized components makes the notional-free scale property very weak in case of technological knowledge. The lack of homogeneous *know-how* capabilities makes the imitation process much more expensive than the one described in the traditional view. One of the main implication of such a change in the theoretical perspective is on the feasible ways that an inventor has to appropriate returns from its invention. The debate on this matter would be unnecessary if we considered knowledge as an immediately and freely reproducible good. In this situation the incentives to undertake an expensive innovative process (that may end with a failure) would be very scarce and limited in time. This is the reasoning justifying the broad usage of strong regimes of patents in the classical standpoints. A deep discussion about strengths and weakness of IPR systems will be the main topic of the next chapter. However, at the time being, it is worth underlining that others tools (beyond patents) can safeguard the inventor's remuneration when it is recognised that knowledge is not a pure public good.

In particular, the means of appropriability typically considered in the literature[25] are:

- Lead time
- Secrecy
- Learning curves

Lead time can have different meanings depending on the contest. When it referred to a possible mean of appropriability it can be defined as *being first to enter the market with a new product and/or being ahead of their rivals* (Laukkanen and Puumalainen, 2007). Carow et al. (2004) argue that first movers advantages are rare, valuable, non-imitable and non-substitutable firstly because these lead to a technological leadership respect to peers.

The idea of secrecy as a tool for appropriating returns simply consists of keeping secret an invention without disclosing it. The main benefits of using secrecy are, intuitively, the absence of fees and the potential no time

[25]See Choen, Nelson and Walsh (2000) and

limitations. On the contrary, it may cause some duplicative innovations that would not occur in case of disclosure.

Learning curves advantages are caused by an increase in learning by doing that can lead to a superiority in technological knowledge (Arrow, 1962). In fact, given the importance of the procedures in the innovative process, being able to apply some tools can be even more important than the tool itself.

The relative weigh of one mean instead of an other cannot be assigned without considering the most common mechanism of appropriation: patent. Therefore, a deeper comparative analysis will be later carried out, after having outline a variety of economic and historical features of IPRs regime. However, at a very general level of discussion, it is possible to state that the best mechanism is not univocally definable. On the contrary, the specific peculiarities (degree of *tacitness*, *cumulativness* etc...) of different sectors play a crucial role.

Chapter 2

Features of the IPR system

Among the variety of mechanisms for appropriation, patents are the most appreciated in the traditional view. Scholars supporting the necessity of strong Intellectual Property Rights (IPR) regimes, indeed, see in the modern definition of patents all the ingredients justifying the spread use of these tools of protection. In fact, the United States Patent Office (USPTO) defines a modern patent as

A legal instruments intended to encourage innovation by providing a limited monopoly to the inventor (or their assignee) in return for the disclosure of the invention[1]

The trade off between the loss of efficiency due to the monopoly provision and the advantages deriving from knowledge disclosure is one of the main argument animating the current debate on patent systems.

In this section we will analyse the general features characterizing the IPR regimes. In particular, attention will be focused on the U.S. case since it represents the largest and the strongest system in terms of norms protecting intellectual innovation.

The aim is, firstly, to underline both historical and theoretical aspects able to describe how the American patent system worked, works and should work. Secondly, some criticisms linked to this structure will be mentioned.

[1] http://www.uspto.gov/web/menu/intro.html Article 1, Section 8 of the United States Constitution

2.1 A brief history of the U.S. patent law

The History of patents began a lot of ages ago. It has been argued that some forms of intellectual monopolies have been existed since the Roman Empire era[2]. However, the first form of an institutional patent system originated in Italy. In 1474, indeed, the first patent statute was issued in Venice. In this document it is actually possible to identify the majority of the concepts that are at the bases of the current IPR regime[3]. The statute introduces concepts of novelty, registration of the new device, term of exclusive right as well as the idea of infringement. In the next two centuries patents did spread across Europe and only in 1790 the first United Stated patent act (entitled *An Act to promote the progress of useful Arts*) was promoted[4].

In 1793[5] the act has been replaced by a more complete e detailed document where, for the first time in the history of patents, the definition of what constitutes patentable subject appears. Moreover, another important feature of the law was that one patent may have a dominating effect over an other.

The possibility of an award of treble damages for patent infringement was introduced in 1800. The entire XIX century has been characterized by both frequent modifications of the law and important Court decisions that contributed to shape the mechanism of the patent system itself. One of the main important changes at a law level was introduced in 1836 with the introduction of the Patent Office in order to ensure the novelty of the invention[6]. It was recognised by the Office after having analysed the claims (never mentioned before) proposed by the inventors. An other important article of the law introduced the possibility of a seven year extension (on the thirteen originally foreseen) of the protection. Simultaneously, the Supreme court introduced the concept of non-obviousness with the Hotchkiss v. Greenwood case [7].

[2]See Devaiah (1992)

[3]Venetian Statute on Industrial Brevets, Venice (1474), Primary Sources on Copyright (1450-1900), eds L. Bently & M. Kretschmer, www.copyrighthistory.org

[4]Copyright Act, Article 1, Section 8, clause 8, New York (1790), Primary Sources on Copyright (1450-1900), eds L. Bently & M. Kretschmer, www.copyrighthistory.org

[5]Patent Act (1793)

[6]Patent Act (1836)

[7]Hotchkiss v. Greenwood; 52 U.S. 248 (1850), www.supreme.justia.com

The two decades 1870s and 1880s have been characterized by the creation of many international IPR organizations and by a gradual recognition of the right of patenting to not American citizens.

This dynamic and rich wave of patent law incrementation partly stopped in 1930s and 40s. These were the years of the Great Depression and of the II Ward War and the supreme court did not appear to be very sympathetic to patents. This aspect is evident, for example, in the Cuno Engineering Corp. v. Automatic Devices Corp[8]. Here the court established that to be patentable an invention *must reveal the flash of creative genius, not merely the skill of the calling.*

After this period of lower consideration and appreciation for patents, the *era of IPR loving reforms* began. The last two decades of the twentieth century saw an increasing number of changes in the U.S. patent policy. The major improvements that have to mentioned are at least four:

- The creation of the **Court of Appeals for the Federal Circuit (CAFC)**. It has been created in 1982 and it is the result of a merge between the United States Court of Customs and Patent Appeals and the United States Court of Claims. In practice, the CAFC institution has been seen a strengthening of the intellectual law since it made easier for patent holders to win legal suits[9].

- The amendment of the **Bayh-Dole Act**[10]. It occurred in 1980 and allowed to universities and other non-profit institutions to get patents from federally funded R&D. In this way, universities are encouraged to collaborate with commercial concerns to promote the utilization of inventions arising from federal funding.

- The expansion of the **realm of patentability**. By a shared decision of the Patent Office and the CAFC the range of patentable subject has been largely broadened in the past two decades. In particular, nowa-

[8]Cuno Engineering Corp. v. Automatic Devices Corp 314 U.S. 84 (1941), www.supreme.justia.com

[9]See Jaffe and Lerner (2004)

[10]P.L. 96-517, Patent and Trademark Act Amendments of 1980

days it is possible to get property right on *invention* as genetically engineered bacteria, genetically altered mice, gene sequences, surgical methods, computer software, financial product and method for conducting auctions on the internet[11].

- The **Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS)**. This is an international agreement reached in 1994 and administrated by the Word Trade Organization (WTO). The agreement establishes the minimum standards of protection to be provided by each WTO member. In practice, this can be seen as an attempt to increase harmonization of the international IPR law[12].

These legal changes have been ascribed to play a crucial role for the explosion of the numbers of patents registered after 1980. However, the evidence on this matter is mixed at the best. The discussion of this issue will be tackled in the next section of this work.

2.2 Theoretical framework

2.2.1 Incentive for innovation

The most common idea linked to the concept of patents is that these are a necessary reward for inventive activities. Therefore, the implicit consideration is that, in absence of intellectual property rights, there would be a lower level of innovation.

Before trying to understand whether it is true or not, it is important to underline that the starting point of this conventional view consists of considering knowledge as a pure public good. If it was so, the first and immediate consequence would be that, with no patents, all the knowledge produced by the inventor would be freely available and usable by anyone

[11] For an exhaustive discussion on this matter see Jaffe (2000)

[12] The impact of the TRIPS as been studied mostly with concern to the possible effects on developing countries. Although this work is not going to analyse this specific issue, there is a wide set of literature dealing with this topic. See for example Cimoli et al(2014),Stiglitz (2008) and Coriat et al.(2006)

else. In this case, the strong positive externality to research becomes cause of a market failure. According to the IPR supporters, a solution consists of making knowledge partially excludable by means of patents; it is equivalent to grant a (limited) monopoly.

The first authors that formally modelled this intuition is Nordhaus (1969). He explained the necessity of a trade-off between static and dynamic considerations in designing patent policy; in other words, he brought to the light the fact that if one wants to incentivize innovation by means of patents, it cannot be done without a loss of efficiency.

The follow-on literature about this theme mostly focuses on how optimally balancing the static costs and the dynamic benefits.

However, it is important to notice that the described trade-off makes sense only under the assumption that there actually exists a positive long run effect associated to patent granting. The latter is, indeed, an assumption and, then, there is not a solid theoretical background supporting it. On the contrary, by looking at the empirical data, the evidence on this relation is ambiguous and sometimes negative, as we shall see in the next section.

2.2.2 Disclosure and diffusion

The importance of a patent as tool for disclosure can be understood when an ex-post dimension in taken into consideration. While the previous argumentations about the incentive to innovate has to be referred to an ex-ante scenario, now it is necessary to focus attention on what happens after the invention has been made.
In the traditional view, the inventor has only two possibilities: keeping the invention secret risking the same discovery is done by another researcher or disclosing it through a patent grant[13].

In the latter case, both the inventor and the society would have a benefit. The first in term of a right of exclusivity and the second in terms of disclosure of knowledge.

Starting from this main consideration, other positive effects are associated

[13]See Gallini (2001)

to the ex-post dimension of granting a patent. Firstly, disclosure could be able to reduce the cost of duplicative R&D or, more generally, it would diminish the double research effort. Secondly, the disclosure of an idea could be source for inspirations of other and new ideas. Finally , after the patent's expiration, knowledge becomes freely available and it would mean a rapid diffusion after 20 years.

Moreover, after the disclosure, the diffusion may come. The general rule is that, following the patent grant, the owner has a total right of exclusivity on the invention. However this is, luckily, not always the case. What leads the IPR supporters to say that patents are good for diffusion is the fact that it is possible to transfer the *covered knowledge* to others, mostly by means of licenses. These confer the right to use the disclosed knowledge without running the risk of incurring in legal suits. It has been argued that the strength of patents plays a crucial role for the amount of given licenses. Here the notion of strength has to be intended as the ease to defend the property right (i.e. the stronger, the easier to defend in case of infringements).

Arora and Fosfuri (2000) showed that the practice of licensing in higher in those industries where the level of effective protection is higher. In this sense, patents can be considered to be self-correcting: the stronger the right to exclude others , the higher the incentive to give and ask for a license. However, despite the relevance of this finding, it does not imply that, in case of a strong IPR system, technology will diffuse more rapidly. Under weak protection, indeed, disclosed inventions can diffuse easily through noninfringing imitations [14].

2.3 Criticisms of IPR regimes

The entire previous section focused on the theoretical framework that has been used in order to explain why a strong IPR regime reveals to be necessary for spurring the innovative process. However, there is a set of alternative possibilities that, if true, would lead to a totally different conclusion. Here,

[14]For further discussion on this matter, see Gallini and Winter (1985)

it is provided an explanation of the theoretical background behind these different views and a review of the main empirical evidences will be helpful in understanding the likelihood for both of the streams.

2.3.1 An alternative standpoint

The nature of knowledge

The basic and unavoidable hypothesis behind the entire classical theory of IPR regimes considers knowledge as a pure public good. Although the two fundamental characteristics that make a good public (non-rivality and non-exclusibility) well apply to the concept of knowledge, its tacit component makes knowledge difficult to be immediately replicated. Under this perspective, a world without a strong IPR system would not necessary mean absence of appropriability. In fact, because of costs, time and personal capabilities necessary for both duplication and learning, the inventor would be able to appropriate some returns even with other mechanisms, alternative to the one built on patents. The immediate and essential implication is that the monopoly argumentations supporting IPR system loss part of their importance if the reduction of efficiency due to the monopoly cannot be considered the only possibility.

Motivation vs external incentives

The trade-off between static costs and dynamic benefits makes sense only in case of real and strong existence of those benefits. Since the positive effect is represented by the power given to patents of generating new knowledge, there is, at this point, a natural question which needs to be answered: *Is it doubtless true that awareness of granting a property right spurs researchers to innovate?*

The most obvious objection that can be moved is that scientists used to do research long before patents existed. This would be not fully justifiable without introducing something more in the standard theory.

Murdock (2002) and Stern (2004) studied the role of *intrinsic motivation*.

In particular, Stern exploited the fact that, prior to accepting a specific job, many professionals receive multiple job offers. Each job offer is composed by a wage and several characteristics. Among them of fundamental importance is whether the employer either gives or not the possibility to keep being part of the academic world, allowing for publications on scientific journals. He conducted an experiment on a set PhD biologists facing different job offers. There are two crucial conclusions deriving from his study. Firstly, as a general trend it holds that the higher the level of freedom (in terms of possibilities to be part of the scientific life) given to the researchers, the lower the wage. Secondly, scientists often have a *science taste*; there are not isolated cases in which they eventually decide to accept the least remunerative job to preserve the possibility of having a research agenda.

This result should not come as a surprise when thinking about the reasons why scientists used to do research long before patent existed. Motivations that incentivize researchers have been the main topic of several studies in the literature, sociological, more than economic (see for example Deci and Koestner, 2001). There is a general consensus in stating that, not only personal and intrinsic motivations play a crucial role in the innovative process, but a system of external rewards may even show negative effects on the amount of effort that people are willing to put in their studies [15]. Now, it might be thought this has little to do with the kind of research that is carried out by those firms that invest in R&D and apply for patents. However, it has to be stressed that technological innovation is highly dependent on a variety of complementary institutions as universities and public agencies; these hardly can be seen as structures where the pure markets incentives are the primary push for innovation.

The fact that patents are not the first determinant for the speed of technological progress seems to be confirmed also from an empirical point of view. Some specific examples on this matter will be later proposed. In general, it has been observed the lack of a positive relation between the strength of the intellectual protection and the rate of innovation [16].

[15] See Deci, Koestner and Ryan (2001)

[16] See Mazzoleni and Nelson (1998); Jaffe (2000); Granstrand (1999); Dosi, Marengo and

Knowledge is the main input for Knowledge

The intrinsic motivation and the strong personal interest represent only a part of the more complex story behind this ambiguous relation. The other (and perhaps the most important) aspects that have to be considered, are the strong *path dependence* and the high level of *cumulativeness* characterizing innovative and technological processes. The entire set of present inventions would not exist without the past research progress, in which both successes and failures have to be included. Therefore, the deeper the awareness about previous research is, the higher the probability of reach a success today becomes. In such a scenario, it becomes reasonable to link the rate of innovation to the level of available opportunities, instead of establishing a positive connection with appropriability. In order to make the things clearer, it is convenient to imagine an *opportunity sea* where incumbents and entrants go fishing for innovation [17]. A broader sea contains more opportunities and these mainly come from the search effort undertaken in the past. In this context, patents are actually seen as responsible for reducing the flow of knowledge present in the sea. Stiglitz (2013c) summarizes this concept stating that *Tigher IPR regimes enable inventors to contribute less to the pool (of opportunities), so that the size of the pool is diminished, so much so that the actual level of innovation may be diminished.*

While the rate of innovation can be metaphorically seen as the *rate of fishing*, the rate of success strictly depends on the firm-specific capabilities (Nelson and Winter (1982); Dosi, Nelson and Winter (2000)). Within this capability-based theory of the firms, it becomes evident that an increasing level of appropriability can hardly have significant effects on opportunities and capabilities, which are among the most important determinants for innovation.

Pasquali (2006)

[17] See Cimoli, Dosi and Stiglitz (2009)

The tragedy of anti-commons

Heller and Eisenberg (1998) stressed the idea that strong IPR systems, not only are unable to increase the rate of innovation, but these would be also dangerous for the entire technological progress. The authors point out the concept of *Tragedy of the anti-commons* as a response to the classical notion of *Tragedy of the commons*. The latter is referred to a situation where individuals, acting independently, use a public good seeking their own interests behaving in a way that goes against the group interest. It can cause the deterioration of the common good due to its excessive utilization. The *Tragedy of the anti-commons* proposed by Heller and Eisenberg, is exactly the opposite. They argue that, when there are too many property rights on a good, a phenomenon of *under-utilization* is very likely to occur. This would be the case of knowledge when too many patent owners are present. The right of someone will very plausibly impede to someone else of developing a new innovation that needs the different, but complementary piece of knowledge covered by the patent. The extreme expression of this problem is represented by the so called *patent thicket*, a situation where the property rights become dense and overlapping. James Bessen has deeply analysed this topic during his research activities. He argues that the probability of thickets is higher in case of complex technologies developed in sector were the patent standards are lower[18]. The patentability standards become a fundamental variable of his model, where patenting itself is seen as economic activity. In fact, the author considers the necessity of trading off the probability of winning a litigation against the cost of patenting; the latter rises when the standards are higher. Bessen concludes that when it is relatively easy to get a patent on a component of a complex product, socially wasteful behaviours are stimulated and the incentives to invest in R&D drastically diminish. Therefore, the idea that stronger IPR systems imply a lower rate of innovation is supported also in this case; the argumentations that lead to this conclusion, however, consider a strategic dimension (chosen by firms but stimulated by strong patentability standards) which has never been mentioned so far.

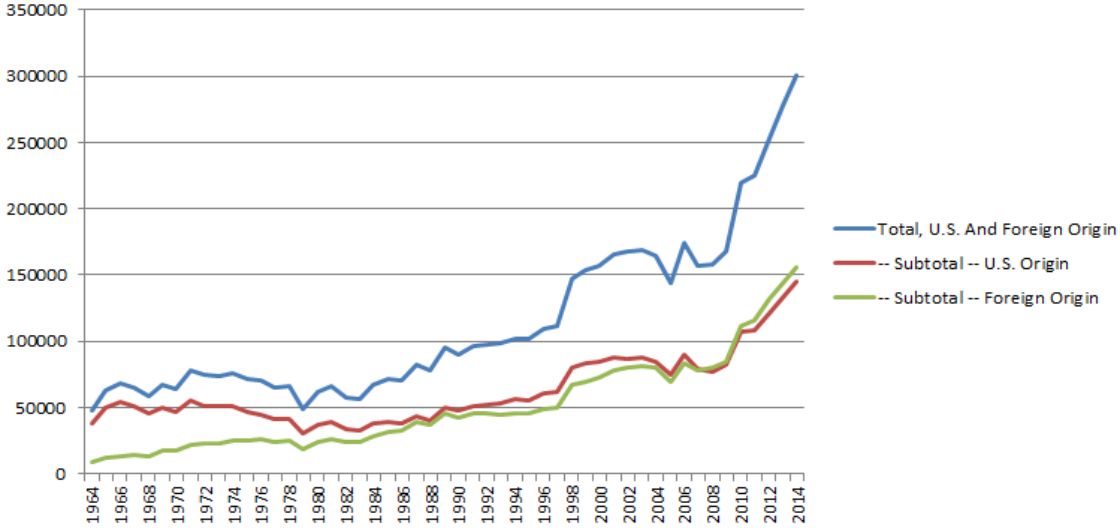
[18]See Bessen (2003)

About disclosure

The cumulativeness of the innovative and technological progress, makes the role of disclosure given to patents of crucial importance. However, the effective capability of patents in disclosing new useful knowledge has been rarely debated in the literature. One of the few authors actively involved in this topic is Fromer. In his paper (Fromer, 2008), the concern is about the uses of disclosure for inventing around, improving upon and inspiring both during and after the patent term. As a result, the author identifies three points of inadequacy in fulfilling the disclosure function that systematically and negatively affect the patent disclosure. The three identified weak-spots are the following. Firstly, because of legal rules of the system, the patent document is poorly structured and it does not contain some of the most important technical information. This aspect incentives the inventor to under-divulge. Secondly, the vast number of issued patents and the insufficient attention to indexing patents makes them hard to be found. Thirdly, a flaw in the disclosure failure is partly due to the behaviours of the readers. The current legal system actually incentives not to read patent to avoid willful infringement. In fact, a court has the authority to award up to treble damages to a patentee when it finds an infringer to have acted willfully[19].
Fromer (2008) proposed some modifications to the current IPR regulation that, in his view would correct the above explained drawbacks. However, when this kind of analysis is carried out in a broader scenario, the suspect that, even with radical changes in the legal environment, some defects still persist is more than plausible. The hight degree tacitness of knowledge, indeed, would hardly make possible to codify all the ingredients, even in case of a *perfect knowledge disclosure*.

[19] 35 U.S.C. § 284 (2000); Read Corp. v. Portec, Inc., 970 F.2d 816, 826 (Fed. Cir. 1992)

Figure 2.1: Number of granted patents by USPTO, 1966:2014



Source: USPTO (2014)

2.3.2 Empirical evidence

From an empirical point of view the first unquestionable evidence is that the number of patents granted by the USPTO[20] has rapidly and drastically grown in the last 30 years. The speed of growth became particularly hight after 1980. The total number of patents granted in 2014 (300678) is almost five times the number of patent granted in 1980 (61819). The rate of increase is, on average, of 5,27% (See figure 2.1). The motivations for such an explosion in the number of patents have been debated at length.

Korum and Lerner (1998) proposed four different hypothesis theoretically able to explain the patent boost. The *friendly court* hypothesis idealizes an higher value associated to patent due to the creation of the CARF. The *regulatory capture* hypothesis, instead, sees the big American firms, reacting to policy changes, as the responsible. Finally, according to the *fertile technology* and the *R&D productiveness* hypotheses, the causes of the explosion would respectively be either an higher rate of new technology or an increasing level of applied research in the innovation processes. In a following stage, the authors tested their hypotheses and they could not find any empirical evidence

[20]Undated States Patent and Trademark Office

that changes in the patent law were responsible for the patent bust. A further support against the *friendly court* and the *regulatory capture* hypothesis has been given in a survey by Choen et al.(2000). Here it is clarified that R&D managers did not perceive patents to be more effective in the 1980s than before. The alternative, largely shared, justification which the authors proposed, is linked with a change in the managers behaviours, who began to use patent as the main strategic tool to block patents of competitors.

Patenting and Innovation

Besides the analyses of the causes for the exponential increasing in the amount of patents, a number of researchers has attempted to test the connection that occurs between patenting and innovation trying to determine not only the existence of a relationship, but also the direction of causality, which is surely an harder task.

Kanwar and Evanson (2003) used cross-country panel data on R&D/GDP investment ratio, patent protection and other country-specific characteristics spanning the period 1981-1995 to directly analyse the relationship between innovation and IPR. In their analysis, the evidence unambiguously indicates the significance of intellectual property rights as incentive for spurring innovation. However, since R&D/GDP is highly correlated with other aspects of the development process, it is unclear whether the measured effect of IPR on R&D intensity is contaminated by causality running from stage of development to strength of IPR[21]. The fact that Kanwar and Evanson (2003) are actually omitting something in their model seems to be confirmed by a set of other empirical studies conducted both on historical and cross-sectional data.

Moser (2005), in her seminal paper, introduces a dataset of almost fifteen thousand inventions at the Crystal Palace Word's Fair of 1851 and at the Centennial Exhibition in 1876 to determine the effect of IPR both on the number and on the direction of innovations. Starting from the evidence that different countries have different levels of effectiveness (See Table 2.1)

[21] Cimoli et al. *Intellectual Property Rights*, p.94

Table 2.1: Differences in patent law accross countries

	Patent length		Population		GDP		Primary education	
Country	1851	1876	1851	1876	1851	1876	1851	1876
Austria	15	15	3,950	4,730	6,563	9,395	389	426
Bavaria	15	—	4,521	—	6,673	—	—	—
Belgium	15	20	4,449	5,303	8,042	14,849	549	582
Britain	14	14	25,601	30,662	60,479	107,661	555	680
Denmark	0	5	1,499	1,973	2,549	4,008	—	—
France	15	15	36,350	38,221	60,685	84,014	515	737
Germany	—	15	—	24,023	—	—	—	732
Netherlands	15	0	3,095	3,822	5,844	52,805	541	639
Prussia	12	—	16,331	—	24,105	—	730	—
Saxony	12	—	1,894	—	2,796	—	—	—
Norway & Sweden	15	—	4,875	—	5,993	—	615	—
Norway	—	3	—	1,803	—	2,650	—	658
Sweden	—	3	—	4,363	—	8,006	—	568
Switzerland	0	0	2,379	2,750	1,986	5,787	—	759
Württemberg	10	—	1,745	—	2,575	—	—	—

Source:Moser (2005)

in patent protection, she found that patent laws shape the direction, but not the rate of innovation. In particular, Moser (2005) argues that in countries with a lower level of protections, inventors tend to focus on a small set of inventions where other legal instrument are more effective than IPR. On the other hand, in countries with a stronger patent law, the inventive activity appear to be more diversified. Lerner (2002) followed a similar approach, analysing causes and effects of differences in the strength of patent protection across sixty countries and a 150-year period. He found the interesting result that the rate of the innovation activity is not significantly influenced by variation in the IPR neither across countries nor in the length of the considered period.

An other wide set of studies tackles the issue of the connection between IPR and patenting focusing on differences across industries, rather then considering cross-countries and time-series data. Levin et al. (1987) conducted this kind of analysis proposing, as a starting point, a questionnaire to a set of high level R&D managers, each one involved in a diverse business. The questionnaire concerned the notions of appopriability, technological opportunities and technological advantages. The collected data show that only

in drugs and petroleum refining industries, patents are seen as the most effective tool for appropriation. In all the other considered cases alternative mechanisms are considered to be more effective. As a consequence, they have found that the impact of legal protection on innovation depends itself on different appropriability conditions which widely differ across industries. Therefore, a general upgrade of the IPR strength cannot be seen as the best solution when it does not consider peculiarities of specific markets. Similar results have been reached also in Cohen, Nelson, and Walsh (2000) where patents are not reported to be the key means to appropriate returns from innovations in most industries.

Moving at the lowest level of abstraction, a number of authors used data referring to a specific industry in order to trace how the evolution of the inventive process has been influenced by IP regimes. Hall and Ziedonis (2001), for example, deeply discussed the U.S. semiconductor industry case. The propensity of semiconductors firms to patent has dramatically risen starting from the mid 1980s, but, at the same this sector is one of those that does not consider patenting as the first source of appropriability and incentives for innovation. This *patent paradox* has been explained by the authors as the result of a practice of *patent portfolio races* which became quite strong in this sector. According to this view, the increased number of patents has been more dangerous than beneficial for semiconductors firms. Similar conclusions have been reached also for the software industry that has been at length studied by Bessan and Maskin (2000). They argue that society and even inventors themselves may be better off without the current strong protection. It is not a chance that both in Hall and Ziedonis (2001) and in Bessan and Maskin (2000), the sectors of interest are characterized by inventive processes which are particularly sequential and complementary.
An additional interesting study deals with the case of steam engines industry. On this matter, Nuvolari (2004) and Boldrin et al.(2008) agree in stating that the diffusion of patents on new engine's component delayed the invention of more efficient machineries.

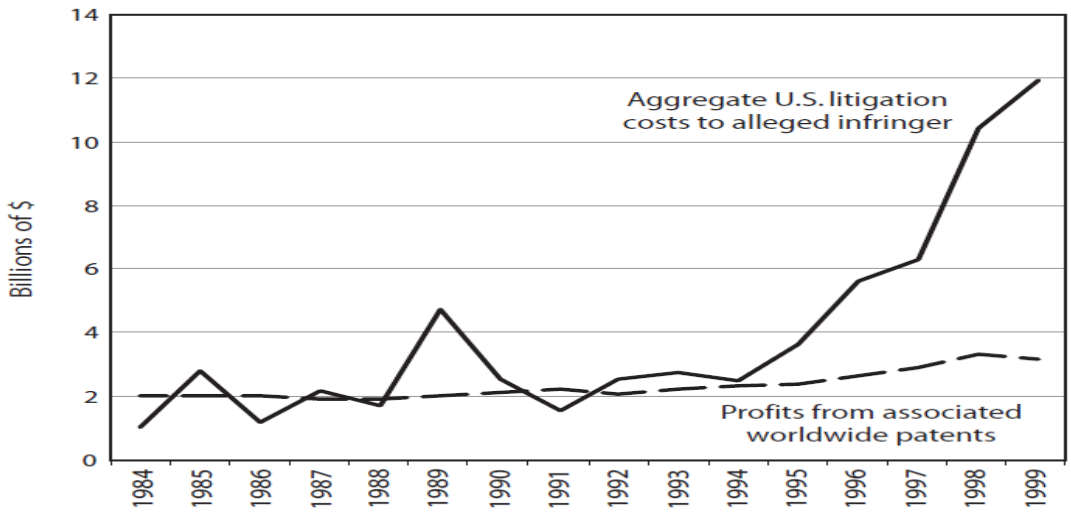
Beyond the tragedy of anti-commons: Troll in action

The problem of the *under-utilization* of commons factor in case of over protection has been only considered as a theoretical warring stressed by Heller and Eisenberg (1998). However, Heller (1998), in his book, goes much further than this. He reported a variety of examples, data and evidences showing that the tragedy of anti-commons is actually much more than an hypothesis. He explains that when too many people own one thing, whether a physical or intellectual resource, cooperation breaks down and everyone looses. Among the most interesting examples, the author reports the case of a group of scientists who had found a potential treatment for Alzheimer's disease, but they could not develop it for the market unless their company bought access to dozens of patents. Any single patent owner could demand a huge payoff; some blocked the whole deal[22]. This example shares a common cause with the wide set of other puzzles proposed by Heller: the over-fragmentation of properties brings to a wasteful under-use of the good. It is worth noticing that here there is neither a problem of appropriability nor a matter of hidden knowledge. On the contrary, the criticism is that, even though inventive advantages have already been available, it become difficult to commercialize it since the costs of sharing (principally due to infringements and legal suits) would be consistently higher than the benefits deriving from the invention. In the last years, the problem has reached its peak when firms have begun to use anti-competitive practices made possible by a systematic failure in the current IPR regime. The most dangerous result of this behaviour is, as already specified, the increasing patent thicket phenomenon. Also on this matter a broad empirical evidence is available and it has been mostly provided by Bessen and Meurer in their book

Figure 2.2 shows one of the most alarming evidence regarding the above described phenomenon. Starting from the middle of 1990s, U.S.firms (different from Pharmaceuticals and chemicals) have seen an explosion in their litigation costs that dramatically reduced their profit. It directly brings the

[22]Heller, *The Gridlock Economy: How Too Much Ownership Wrecks Markets, Stops Innovation, and Costs Lives*,NewYork: Basic Books, 2008, Preface

Figure 2.2: Aggregate profits from patents and aggregate litigation costs for U.S. public firms (Pharmaceuticals and chemical firms excluded)



Source: Bessen and Meurer (2008)

authors to the strong and hardly debatable conclusion that, including the risk of litigation, a number of industries (different from chemical and pharmaceuticals), would be better off if patents did not exist[23].

Furthermore, the problem related with litigations costs lowering both the degree of innovativeness ant firms' profits, has been exacerbated by the recent diffusion of the called *patent trolls*. Tucker (2013) defines them as classes of patent owners who acquire patents from inventors but who do not intend to provide products or services themselves. On they contrary, they just use them as a legal tool to impede the works of others who cannot use the invention disclosed in a patents without paying realities to the troll.

[23]Bessen and Meurer, *Patent Failure:How Judges, Bureaucrats, and Lawyers Put Innovators at Risk*, 2008, p.16

Chapter 3

Pharmaceutical and medical innovation

Pharmaceutical industry has always been considered as an R&D intensive and science-based industry where innovation is driven simultaneously by research in scientific and applied knowledge. The interesting feature of this industry is that starting from the mid-1970s, it underwent a paradigmatic technical change[1] that fundamentally changed the research process and dramatically modified the entire industrial structure.

At the beginning, pharmaceutical innovation was almost totally dependent upon an oligopoly of large firms that were involved both in the innovative and in the commercial activity. Without going in depth through the explanation of the industrial evolution, there is a general consensus in stating that before the mid-1970s only a small number of firms entered the pharmaceutical industry (Dosi and Mazzucato, 2006).

Since then, the emergence of small biotechnology firms became a growing phenomenon. It has been argued that this change in the pharmaceutical industrial structure can be associated to a shift in the paradigm underlying research activity. In particular, Gambardella (1995) described two different technological paradigms dominating respectively before and after 1970.

In the first period search activity was mostly based on *random screening*.

[1] See Gambardella (1995)

In the second period, instead, research process became more *guided* and more based on theoretical advantages in chemistry and molecular biology. A main implication of the paradigmatic change is a shift in the technological regime with an higher level of *path dependency* and *cumulativeness* of knowledge.

One of the most important consequences of the *molecular revolution* has been the division of innovative labour (Arora and Gambardella, 1994). In the new scenario, indeed, small biotechnology firms, publicly founded organizations and universities focus on upstream research where the level of cumulativeness is much higher, while pharmaceutical firms, that are still large and relatively few (Big-Pharma), cover that part of the process related to the trials.

Therefore, when the aim is to consider potential effects of any factor on pharmaceutical industry, it is necessary to keep in mind the complementarity between these two separated, but strictly interconnected, areas. Hence, the first section of this chapter will be dealing with issues referring to IPR in big-pharma. Later, a discussion about patents granted to universities and other public institutions will be carried out.

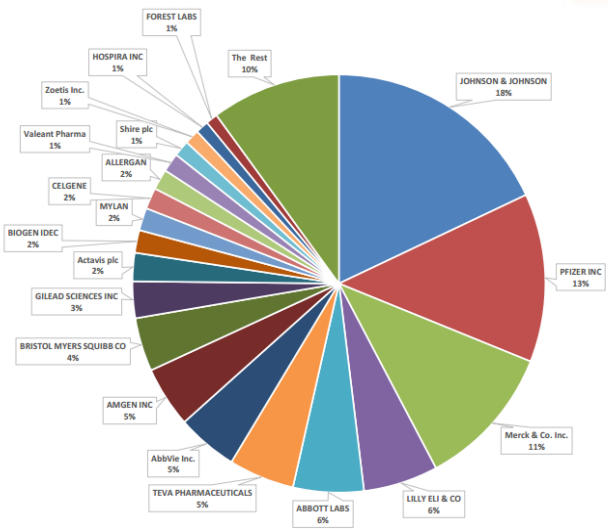
3.1 Innovation in Big-Pharma

Finding detailed data about costs for R&D spent by pharmaceutical firms is not an easy task, because of their scarce propensity to share it. However, by looking at some statistics conducted at the aggregate level, it is possible to identify some trends in the sector that are, at the best, ambiguous.

First of all, it is worth noticing that, in case of pharmaceutical industry, 90% of the market is controlled by no more than 20 firms as shown in Figure 3.1. This evidence is perfectly in line with the premise that, even after the aforementioned shift in paradigm, the structure of the market keeps being characterized by a small number of big firms that make difficult the entrance for new firms.

The broadest set of data about the sector is available in the *annual survey reports* yearly published by the Pharmaceutical Research and Manufacturers of America (*PhRMA*). This association does not represent the whole popula-

Figure 3.1: Division of 2013 revenue across pharmaceutical firms



Source: Calcbench, Pharmaceuticals Industry Analysis

tion of the firms operating in the pharmaceutical sector, however its members account for the 90% of total sales. Therefore, the data disclosed in the reports are a good proxy for general trends occurring into the entire sector of interest.

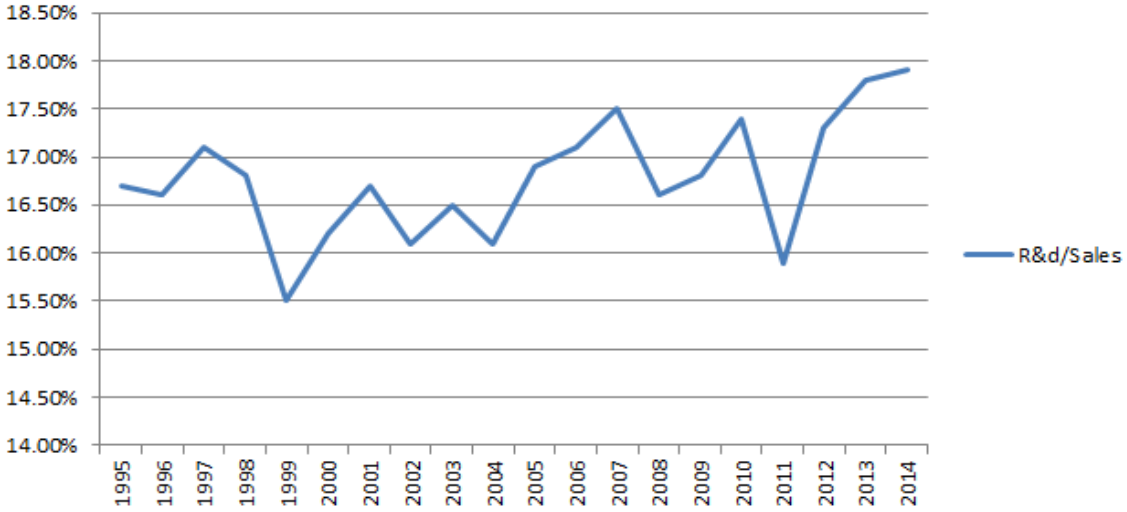
The most relevant aspect that has to be underlined is the high rate of R&D expenditure on total sales. As shown in Figure 3.2 this has never been lower than 15.5% during the last 20 years. Comparing the value to the equivalent in other sectors, it reveals to be much above the average (3.9%)[2]. Intuitively, it can be interpreted as a signal of innovativeness which characterises the main pharmaceutical industries operating in the sector.

However, the output of the R&D investment is very far from being homogeneous (See Figures 3.4 and 3.5). What is unambiguous, instead, is that the investment efforts of pharmaceutical industries have been well remunerated in the last decades[3] . The other two factors that have to be considered when trying to solve the *PhRMA puzzle* are (i)the growing number of patents

[2]See Nicholson (2012)

[3]Firmas belonging to *PhRMA* have a median profit margin of 17%; a measure well above the the other sectors' average which show a value of 3.1% (Pattison and Warren, 2002)

Figure 3.2: R&D-Sales ratio for *PhRMA*



Source: Based on data from PhRMA 2015 annual report

associated with the FDA (Food and Drug Administration) approvals and, simultaneously,(ii) the above average profits registered by the sector.

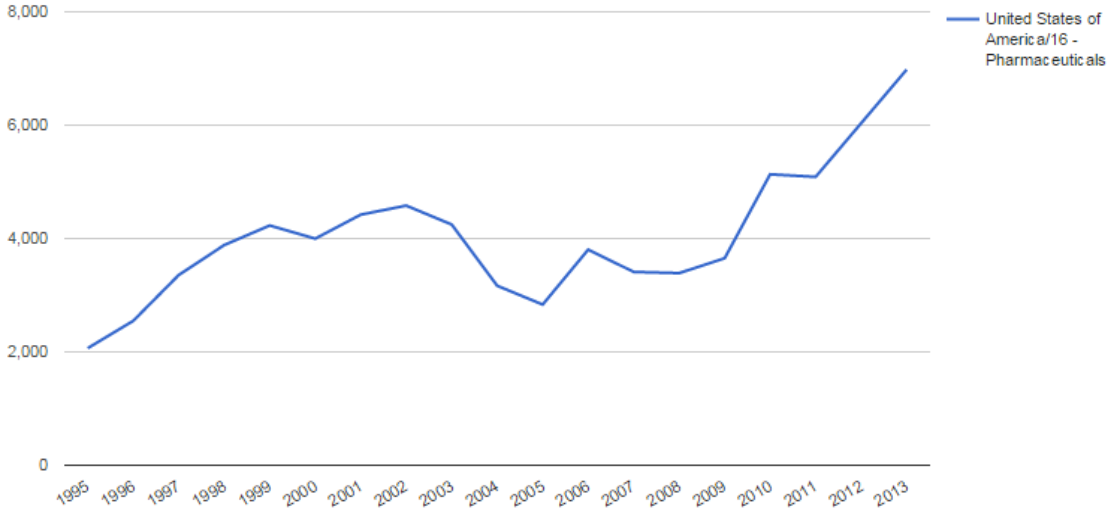
3.1.1 The heterogeneous quality of patents for drugs

Figure 3.3 shows the explosion in the amount of patents granted to pharmaceutical industries from 1995 to 2013. As already clarified in section 2.3.2, it is difficult to associate an increase in the IPRs with an equivalent improvement in the rate of innovativeness. However, the pharmaceutical case is a peculiar one, since a clear classification of approved drugs helps to identify the length of the innovative step behind each grant. Specifically, the FDA uses two different criteria in order to categorize new drugs; one in terms of chemical characteristics and the other in a view of therapeutic potential[4]. The first dimension is referred to the compounding forming the active ingredient of the product which can be classified in one of the following classes:

- *New molecular entities* (NMEs) containing active ingredients never approved before.

[4]See www.fda.gov

Figure 3.3: Number of patents in pharmaceutical sector



Source: ipstats.wipo.int

- *Incrementally modified drugs* (IMDs) sharing the same active ingredients with existing drugs, but that, thanks to some modifications, are considered to be safer, more effective or more convenient to use.

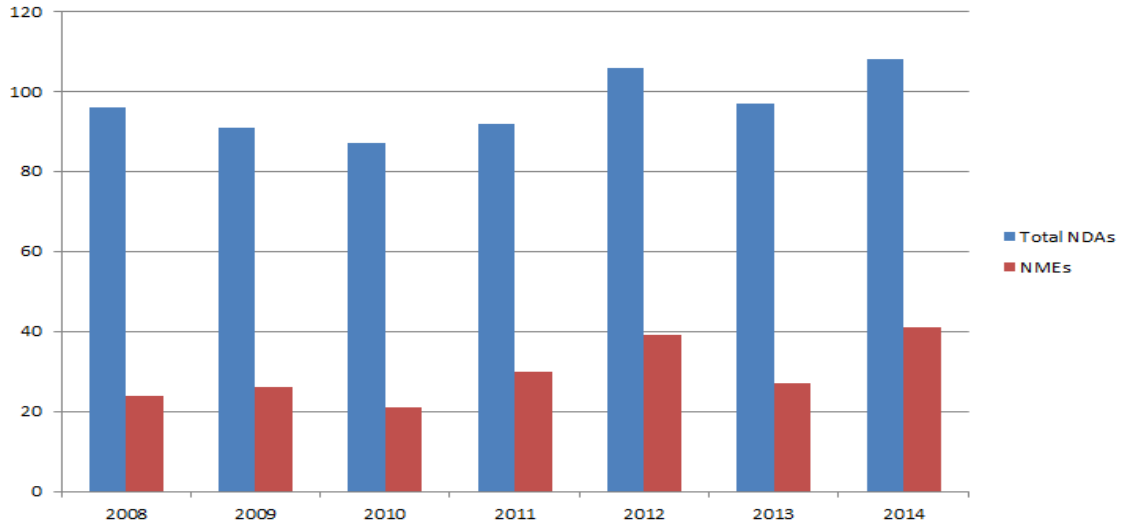
This differentiation clearly allows to distinguish between incremental and true innovations.

When considering the therapeutic potential, instead, drugs can be included by FDA in one of the following categories:

- *Priority review* if the drug can demonstrate even a moderate clinical improvement.
- *Standard review* when the drug falls to demonstrate a clinical advantage.

Thanks to the annual reviews annually published by the FDA it is possible to have an exact measure of the number of new drugs approvals (NDAs) recognised in a specific year. In the NDAs category are included all the typologies of drugs, regardless of the features above mentioned. However, the FDA reviews also furnish data about the product classification both in terms of chemical compounding and therapeutic potential.

Figure 3.4: Share of NMEs on total NDSs



Source: FDA.gov

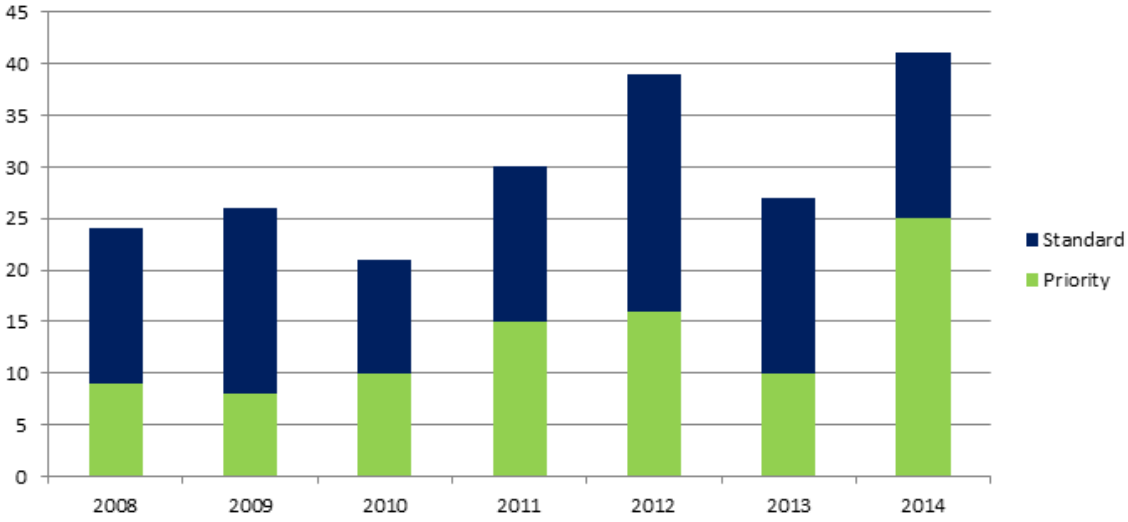
From 2008 to 2014, the FDA approved 677 NDAs. Of these, only 198 had a truly new component. The large majority (469), instead, had active ingredients which were already available in a marketed product. Figure 3.4 shows the annual trend for the new drugs approvals by FDA and underlines the portion of NMEs on the total since 2008 to 2014, highlighting the limited relevance of NMEs.

Moreover, the linkage between the chemical nature and the therapeutic importance of a drug does not have to be taken for granted. In other words, the fact that a drug is innovative in terms of chemical characteristics does not mean that it is also considered important for therapeutic purposes. Figure 3.5 puts in evidence how much of the approved NMEs has been assigned a priority review.

Combining the previous observations, it can be stated that despite the high number of drugs approved by the FDA between 2008 and 2014 (667), less than a third are NMEs (30%) and only 13% of them are recognised to have a priority review (See Figure 3.6).

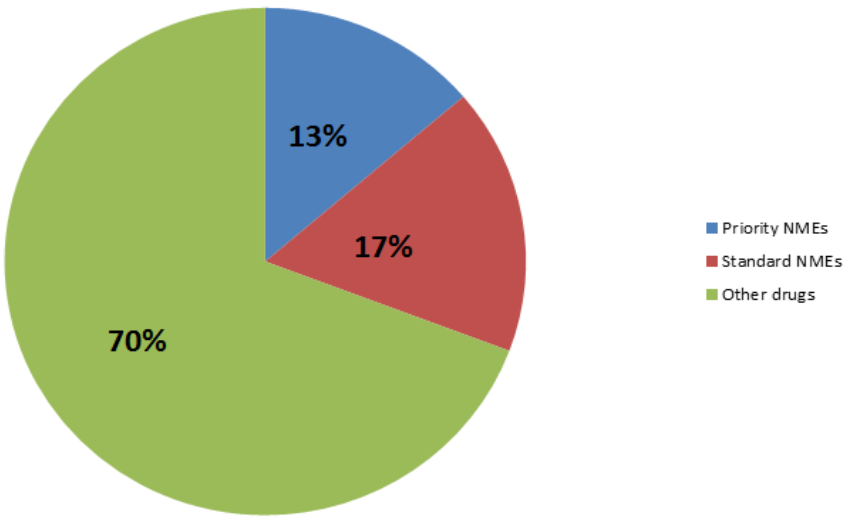
This result, referring to the last seven years, seems to confirm a big warring often pointed out in the literature during the last decades: the limited

Figure 3.5: Portion of priority on standard review for NMEs



Source: FDA.gov

Figure 3.6: Division of NDAs approved between 2008 and 2014



Source: FDA.gov

innovative output of Big Pharma.

Angell (2004) deals with the problem of *me-too drugs*, defined as those ones that present only minor variations of highly profitable pharmaceuticals already on the markets. Specifically, she argues that 78% of the FDA approvals recognised between 1998 and 2003 were actually nothing but the outcome of little modifications of older ones. This estimation coincides with the one proposed in Boldrin and Levine (2008), who argue that more than 77% of FDA approvals are *redundant from the strictly medical point of view.*

Several reasons can lead to the development of a *me-too drug.* First of all, these can be used as a mean through which competitors can gain competitive advantages in the markets. Second, the introduction of drugs only slightly different from other profitable drugs (*blockbuster*[5]) already available, results in the possibility to extend the monopolistic right on older drugs (Angell 2004).

Apart from market reasons motivating the launch of *me-too drugs*, for sure they are cheaper and less risky to develop. In fact, by definition these drugs heavily rely on existing knowledge and competence. However, the IPR regime works in the same way for every type of products regardless of their quality characteristics. Whether a drug is merely replicative or particularly innovative, in both cases a monopolistic right of 20 years is typically recognised.

3.1.2 Profits in Pharmaceutical industry

Although the claims of highly risky activity by pharmaceutical industries, their profits are, on average, higher than any other sector. Pattinson and Warren (2002) estimated that the top 10 drugs companies in U.S. (all of them belonging to *PhRMA*) have a median profit margin of 17%. This feature appears to be particularly substantial when compared to the same value registered in all the other industries present in the Fortune 500 list, that is (on

[5]Blockbuster drugs are defined as those ones which lead an annual profit of at least $1 billion of profit to the company producing them.

average) 3.1% [6]. Such a deep difference must obviously find an explanation in the excessive differential between the cost of production and the drug prices. In fact, granting monopolies on drugs, has a fundamental consequence on product prices. The latter are not subject to any regulation and can be fixed at high level because of the strong IPR regime operating in pharmaceutical sector. Apart from moral considerations on this matter, it would be interesting to unfold the factors that allow the pharmaceutical firms to maintain their above average rates of profit. For doing it, all the aspects mentioned so far have to be considered as fundamental pieces of a complex puzzle. First of all, the peculiarity of high R&D expenditure characterizing the pharmaceutical industry has been object of a number of studies which, somehow, tried to identify the portion of expenditure really devoted to risky and uncertain activities. Love (2003), for example, estimated that only the 2.6% of R&D expenditure is actually devoted to priority review drugs. This aspect leads to infer that the high portion of non-innovative drugs approved by FDA, is the consequence of a strategic choice, rather than the effect of failures linked with a risky and costly innovative process. This evidence seems to contradict an old campaign carried out by *PhRMA* to convince both public and policy makers with the claim that the industry needs extraordinary profits to fund expensive, risky and innovative R&D. In the words of Holmer, the *PhRMA* president:

Believe me, if we impose price controls on the pharmaceutical industry, and if you reduce the R&D that this industry is able to provide, it's going to harm my kids and it's going to harm those millions of other Americans who have life-threatening conditions. (National Public Radio, *Talk of the Nation*, Jan. 2, 2001)

Later, he reinforced the argument claiming that the cost of conducting a single drug to the market, amounts to $500 millions. This estimation is coherent with a study conducted by Di Masi, Hansen and Grabowski (2003)

[6] The estimation of this values is specifically referred to 2002. However, the observation of higher profits in case of pharmaceutical industries appears as a regularity persisting in time.

where the authors considered how each phase of the development process of a drug contributes to increase the cost of production. In fact, the development of new drugs is a multi-stage process. It begins with a preclinical phase[7] and then it passes thorough three different phases of testing on humans. During Phase I, the drug is given to a small set of patients in order to exclude the possibility of causing considerable side effects. Lather, the drug is administered to a broader set of people to test the effectiveness of the product. Lastly, during Phase III the treatment is tested on at least two large samples of patients. Only at the end of these costly (but not R&D intensive) steps, if the product has been recognised safe and efficient enough, the firm can apply for a FDA approval. According to the authors, the span time necessary to reach the end of development process is, on average, 11.8 years and the rate of final approval is equal to 12.5%. By considering these aspects, and recognising that the costs of factors increase during the period, Di Masi et al. concluded that the average cost per approved new drug is $467 millions. This measure has been rounded up to $500 millions by the industry (Manthien, 2003). Despite the relevance of this study the authors, some criticisms[8] have emerged highlighting the possibility that the cost may actually be overestimated. The weakest point is surely linked with the type of drugs considered in the study. Although the industry uses the proxy of $500 millions as the price of any new approved drug, the products involved in the Di Masi et al. seminal paper are all NMEs [9]. Furthermore, the focus is only on *self-originating drugs*, which are new entities developed by the company as opposed to those they acquire from other research organizations. These aspects, together with the observation that the percentage of R&D expenditure devoted to NMEs is lower than 3%, imply that the $500 millions measure plausibly overestimates the real cost of production.

[7] The preclinical phase can actually be divided in discovery and preclinical test costs. Since the availability of data on this matter does not allow for disaggregating them, Di Masi et al (2003) consider the two categories together labelling them preclinical costs

[8] See (Manthien, 2003)

[9] The category of drugs used by Di Masi et al. is, actually, slightly different from NMEs. In fact, the authors refer to new chemical entities (NCEs), instead of new molecular entities. However this is a matter of notation rather than relevant from a substantial point of view.

All the discussion above leads to say that although the high rate of R&D expenditure shown in case of pharmaceutical industry, a positive relation with the degree of innovativeness can hardly be established. Moreover, the claim that *PhRMA* needs extraordinary profit for the development of expensive and risky processes becomes less credible when a broader set of factors is considered. Finally, although the strategic behaviour of investing in remunerative, but low innovative activities intensifies the problem of the absence of controls on drug prices, the policy structure behind pharmaceutical innovation does not seem to aim at a solution. Specifically, one of the most important tools which is essential for the industry profit is represented by the IPR system that, in the specific case of big-pharma, reveals to be essential for approbriability, but not for innovation. One of the biggest danger linked with such an high level of protection recognised for drugs (without distinction among the several categories) is that it risks to stimulate *anti-innovative* activities. When a monopoly is recognised for both *blockbuster* and *me-too* drugs, the choice of producing the latter becomes too easy not to be taken.

3.2 The upstream level of medical innovation

Although in the previous section the concept of *self originating drugs* has been just mentioned, it actually deserves an high level of attention. In fact, Nicholson (2012) proposes a very interesting analysis on the role that public and other institutions play for the medical research process. Specifically, he pointed out that R&D expenditures devoted to medical innovation in U.S. shows a division in terms of sources of funding that can be summarised as follows:

- Public institutions finance 37%[10]
- Pharmaceutical industry account for 34%
- Biotech firms cover 24% of the expenditure[11]

[10]The public institutions considered in this category are NIH, state and local governments and other federal institutions.

[11]the remaining 5% is ascribed to other foundations and private organizations.

This division is one of the effects of the so called Bio-tech revolution that contributed to reshape the industrial structure after 1970. In fact, before the aforementioned shift in paradigm, almost the totality of innovative labour in medical field was conducted by pharmaceutical firms. Nowadays, instead, both the role of public and the diffusion of biotechnological firms have to be considered in order to have a complete idea about the peculiar innovation process in medical research.

3.2.1 Role of public in medical research

The first general element that underlines the importance of public funding for medical research deals with the high percentage (20%) of federal R&D expenditures devoted to medical research. In opposition with the pharmaceutical trend, the majority of this funds are used in basic research rather than in applied. In fact, Nicholson (2012) states that 55% of NIH financing in 2003 targeted the upstream level of research and this proportion kept being stable between 1994 and 2003 (Moses et al 2005). A number of studies have analysed the issue of a possible relation between the NIH and pharmaceutical innovations. Among them, Toole (2008) found out that a 10% increase in NIH research is associated with a 6% increase in new molecular entities[12].

Moreover, when a division between public and proprietary source of knowledge is marked, some universities and laboratories must be included in the first category. Referring to medical research, their contribution for the most innovative steps has rarely been objected. Cockburn and Handerson (1996), basing their analysis on qualitative and quantitative data referring to a set of drugs, found evidence of a significant interaction between public and private spheres of the medical innovative process. Mansfield (1998) argues that 31% of new drugs produced between 1986 and 1994 could not have been developed without the fundamental research conducted by universities. Vallas et al. (2011) reached a similar conclusion after having analysed a sample of fifteen *blockbuster drugs* developed in U.S. during 2006. They found out that only

[12]For additional studies on this matter see, for example, Toole (2007) and Blume-Kohout (2009)

two products belonging to the sample can be described as *self-originating drugs.*

3.2.2 The Bio-Pharma revolution

A Bio-pharma firm is conventionally defined as a small firm that conducts discovery research and has been founded after 1970[13]. In fact, consequently to the shift of technological paradigm (described at the beginning of this chapter), the search activity in the field of medical innovation has largely relied on the fast advantages in the fields of biotechnology and, especially, of genetics. It led to a division of labour between rooted pharmaceutical corporations and young biotech firms.

The changes occurred in the industrial structure in the '70s, has stimulated the interest of a number of scholars. Many of them [14], focused on the financial characteristics of the new firms. Specifically, they tried to identify the main channels of funding operating for the industry. Despite the great importance of the topic, is not the aim of this work to deal with themes concerning the possible constrains that biothec firms have to face for financing their projects. In fact, for the sake of a complete comprehension about the path of medical innovation, it is worth considering the features of the innovative activities conducted by the firms, rather than their *budget problems.* Dosi and Mazzucato (2006) argue that small dedicated biotechnology firms typically concentrate on upstream research developing initial drug compounds later bought by big pharmaceutical industries. This aspect further strengthens the idea that the portion of NMEs developed inside pharmaceutical firms after a risky and a long innovation process is actually very low.

3.2.3 Monopoly rights in basic research

In order to study the possible effects of IPR at the upstream level of medical research, two main factors have to be considered. Firstly, the Bayh-Dole

[13] See Nicholson (2012)

[14] For further discussion on this topic, see Metrick and Nicholson (2009) and Guedj and Scharfstein (2004).

act (1980), which allowed to publicly founded universities and public labs to cover with patents their inventions, signed one of the most important policy changes for the basic research dynamics. Secondly, the field of research mostly exploited by both publicly founded institutions operating in medical research and small biotech firms is, doubtless, human genetics. Therefore, the debate over public and private contributions to genetic knowledge fuelled by the rapid sequencing of the human genome, the burgeoning stream of public genetic knowledge published in scientific articles and the expansion of gene patenting by industry – exemplifies broader arguments over whether patenting helps or hinders public knowledge (Huang and Murray (2009), p.3). In accordance with other scholars, Greenfield (2006) argues that patenting on genetic material is creating numerous limitations to scientific process in the area of medical research. The practice of patenting both full genes and small part of DNA sequences, is increasing gene test price, initiating costly patent wars between firms and universities. On the contrary, Walsh et al. (2005) argue that intellectual properties have actually no effect on public knowledge of genetic.

In order to find a clear answer to the role on monopoly rights in genetic research, several studies have tackle the problem focusing on single cases of litigations and infringements. Merz and Cho (2005) state that among the wide variety of patents protecting genetics knowledge, important ones are those referring to all known methods to test disease. The most famous gene patents dealing with this issue are those implicated in breast and ovarian cancer (BRCA1 and BRCA2) and those referring to the test of the Alzheimer's disease (Apo-E). The problematic aspect related to this kind of patents is that the genes causing the disease can be multiple and leading to the possibility of having a patent thicket of claims overlapping in the genetic territory. This is the case, for example, of patents on BRCA1 and BRCA2 genes, which are at least twelve only in the Undated States. Furthermore, the same gene can be used to test different prognostic purposes, as in the case of Apo-E test. Although the majority of gene patents owners tend to be rather flexible in enforce their exclusivity against clinical laboratories, the latter report to have had at least one gene patent asserted against them (Cho et al. 2003).

Typically, clinical and research laboratories pay royalties for using patented technologies. However, as observed in Merz and Cho (2005), in a rather small number of cases patent owners even refuse to grant licenses preventing to use the knowledge about a certain gene in disease's testing. The case of Myiad Genetics, which owns BRCA1 invention, is exemplar for this matter. In fact, Myriad Genetics has been recently involved in a legal battle where the Association for Molecular Pathology accused the company to held property rights on genes that naturally occur on every human and to obstruct the fundamental research on ovarian cancer. The Supreme Court confirmed this idea declaring that *A naturally occurring DNA segment is a product of nature and not patent eligible merely because it has been isolated*[15] and invalidating two Myriad Genetics' gene patents.

Although these specific examples well describe the challenges of patents on genes, they represent only few cases that can be treated almost anecdotally. In order to develop a large scale study on the interaction between proprietary and public knowledge, instead, Murray (2002), proposes an innovative methodology, based on the construction of *patent-paper-pairs*, that allows for the identification of pieces of knowledge simultaneously disclosed by papers and patents. Since we find particularly powerful this strategy, we decided to rely mostly on Murray's work to conduct our analysis on the potential effect of IPR regimes on the highest level of medical research, Therefore, a deep discussion about *patent-paper-pairs* approach will be carried out in the remaining part of this work.

[15]Supreme court,.No. 12–398, June 13, 2013

Chapter 4

Data & Method

The aim of this work is to investigate the interactions between public and private tools of knowledge disclosure. In particular, our ambition is finding concrete pieces of knowledge that, in different periods of time, have been disclosed both by means of patents and publications. To achieve the goal, we followed a procedure based on Murray's (2002) seminal work, that introduces an innovative approach, based on the construction of *patent-paper-pairs*, able to identify those *touching-points* between public and private spheres of knowledge. We think this is the best procedure to give an answer to our main research questions, which are:

- In which way knowledge disclosed by both patents and publications differs from information disclosed only by means of private tools?
- Is the path of follow on knowledge influenced by the granting of a patent associated to the same piece of knowledge?

4.1 Patent-paper-pairs

A *patent-paper-pair* can be defined as a set of two documents -a patent and a scientific publication- that, despite being subject to different institutional contexts, share exactly the same content. In particular, patents refer to a proprietary dimension of knowledge, whereas papers published in scientific

journals contribute to increment the so called stream of public knowledge. Following Murray (2002), in order identify a *pair* the patent and the paper must show a correspondence between authors, publication dates and contents. The starting point for the couples' identification process can be either a set of papers (Murray, 2002) or a set of patents (Huang and Murray, 2008) depending on the aim that the researcher wants to achieve.

Since we want to focus our a attention on the specific case of innovation about cancer and tumors, we decided to use a set of patents, explicitly referring to this matter, as starting point for our analysis.

Data about patents have been taken from the USPTO FULL DATABASE available on the internet[1]. It contains all the patents issued by the *United States Patent and Trademark Office* (USPTO) since 1976 and it allows to retrieve the data using different criteria such as: assignees, applications, dates etc. For this work, we found particularly useful to search the initial patents set using the field *current U.S. classification* containing the technological class to which the patent is assigned. The one we considered is the number 435/7.23, which contains, according to the USPTO declaration, all the inventions concerning *tumor cells or cancer cells*. The class includes 2282 patents that have been granted in the last 39 years. Our analysis, however, focuses on 761 patents granted between 2004 and 2011. This 761 patents constitute our initial set.

In order to identify which patents of the initial set constitute a *pair*, we used the ISI WEB OF SCIENCE DATABASE[2] which contains all data we need about scientific publications. It is worth noticing that, although the USPTO FULL DATABASE and the ISI WEB OF SCIENCE contain very detailed information respectively about patents and papers, these are totally independent one from the other. Therefore, the construction of the *pairs* has been based on a *manual matching* repeated for each one of the 761 patents contained inside the original set.

The first step of the identification process is related to the people involved in the inventive process. In fact, at the initial stage, we checked for each

[1] www.uspto.gov

[2] The full patent database is available on apps.webofknowledge.com

patent, whether its inventors appear also as authors of some peer reviewed articles[3] contained in the ISI WEB OF SCIENCE. We used a rather restrictive definition according to which all the inventors have to be listed among the authors[4].

In cases where the *author matching* condition was satisfied, we proceeded to compare the patent application date and the publication date, requiring that they do not differ for more than two years. In some cases, however, we found that, despite a considerable temporal dispersion, patents exactly retrace the content of certain papers. In exploring the causes of such apparent contradiction, we recognised a correspondence between the priority date (rather than the application date) and the publication date. This scenario is particularly frequent for patents with non-american inventors that apply for a patent in a different country claiming, simultaneously, a priority in USA. The result is that, although the USPTO DATABASE only shows the date in which the document was filled in U.S., it is actually referred to a piece of knowledge already opened in another application made in a different state. Therefore, we decided to consider suitable for the *date-matching-requirement* also those patents that show a link between the priority date and the publication date[5].

The last condition for the construction of a *patent-paper-pair*, relates to the treated topic. In particular, we checked that the patent abstract matched the paper abstract. It is worth noticing that, because of a lack in specific biological and biomedical knowledge, the *abstract matching* condition has been tested mostly through a keyword searching activity.

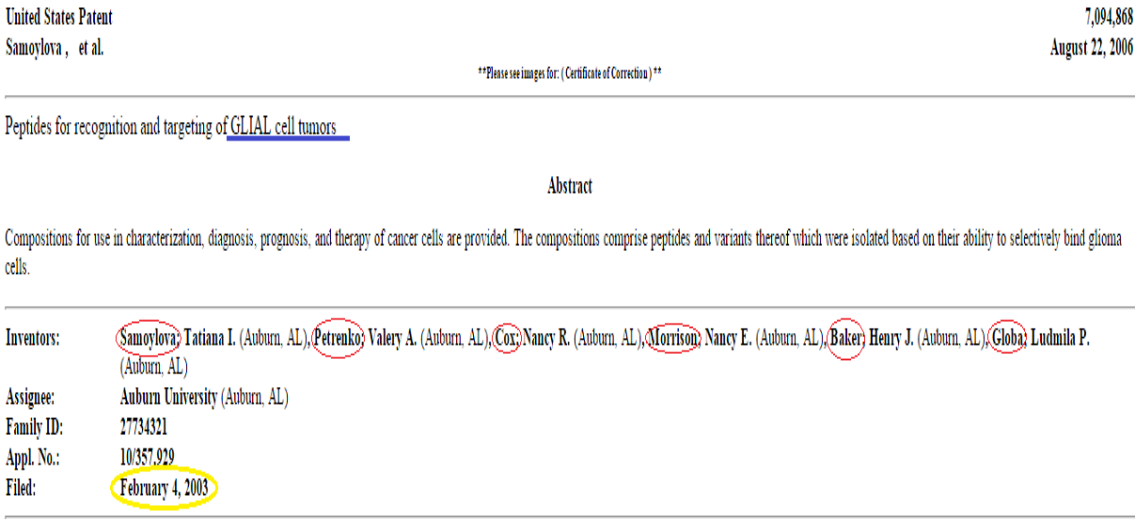
From the initial set of 761 patents, we were able to find, at the end of the matching process, 202 documents that undoubtedly contain pieces of

[3]These refer to those articles that have been submitted to a process used for checking the work performed by one's equals (peers) to ensure it meets specific criteria. The evaluation is, therefore, made by scientific, academic or professional team operating in the same field of the author.

[4]Note that the following relation does not hold. In other words, the paper's authors can be more than the patent's inventors

[5]To give a clear example of this situation, patent number *6,756,201* can be taken into consideration. On the USPTO, it appears that its *application date* is February 2001, however, by conducting a deeper research (on GOOGLE PATENTS), we found out there was a priority date on this document dating back to 1995. Therefore it is worth associating it with a paper published much earlier than 2001.

Figure 4.1: Patent: number 7,094,868



United States Patent 7,094,868
Samoylova , et al. August 22, 2006

**Please see images for: (Certificate of Correction) **

Peptides for recognition and targeting of GLIAL cell tumors

Abstract

Compositions for use in characterization, diagnosis, prognosis, and therapy of cancer cells are provided. The compositions comprise peptides and variants thereof which were isolated based on their ability to selectively bind glioma cells.

Inventors: Samoylova; Tatiana I. (Auburn, AL), Petrenko; Valery A. (Auburn, AL), Cox; Nancy R. (Auburn, AL), Morrison; Nancy E. (Auburn, AL), Baker; Henry J. (Auburn, AL), Globa; Ludmila P. (Auburn, AL)
Assignee: Auburn University (Auburn, AL)
Family ID: 27734321
Appl. No.: 10/357,929
Filed: February 4, 2003

Source: patft.uspto.gov

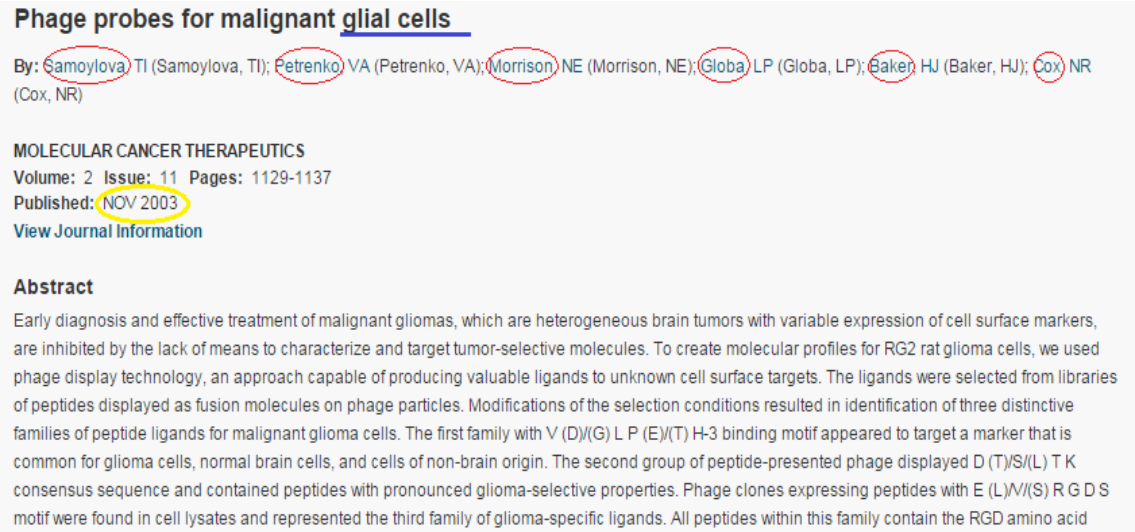
knowledge previously disclosed in a scientific journal. These 202 couples of patents and publications constitute the set of PAIRED-PATENTS studied in this work. The remaining 559 patents not associated with a publication constitute the a set of (NO-PAIRED-PATENTS). Figure 4.2 and Figure 4.1 give a clear example of a *patent-paper pair* that perfectly satisfies the three criteria above described.

Beyond the USPTO FULL DATABASE and the ISI WEB OF SCIENCE we collected data about patent characteristics from the EPO-PATSTAT DATABASE(April, 2014 version) and the OECD QUALITY DATABASE.

4.2 Method

In order to address our research questions we are going to use both descriptive and inferential statistics. First of all, we are going to analyse and compare several characteristics of the PAIRED PATENTS and NO-PAIRED PATENTS, considering them as two independent populations. The comparison cannot takes advantage of the classical statistical tools applicable in case of variables normally distributed. Specifically, the standard *t-test* on mean differences

Figure 4.2: Paper: WOS:000186534300006



Phage probes for malignant glial cells

By: Samoylova TI (Samoylova, TI); Petrenko VA (Petrenko, VA); Morrison NE (Morrison, NE); Globa LP (Globa, LP); Baker HJ (Baker, HJ); Cox NR (Cox, NR)

MOLECULAR CANCER THERAPEUTICS
Volume: 2 Issue: 11 Pages: 1129-1137
Published: NOV 2003
View Journal Information

Abstract

Early diagnosis and effective treatment of malignant gliomas, which are heterogeneous brain tumors with variable expression of cell surface markers, are inhibited by the lack of means to characterize and target tumor-selective molecules. To create molecular profiles for RG2 rat glioma cells, we used phage display technology, an approach capable of producing valuable ligands to unknown cell surface targets. The ligands were selected from libraries of peptides displayed as fusion molecules on phage particles. Modifications of the selection conditions resulted in identification of three distinctive families of peptide ligands for malignant glioma cells. The first family with V (D)/(G) L P (E)/(T) H-3 binding motif appeared to target a marker that is common for glioma cells, normal brain cells, and cells of non-brain origin. The second group of peptide-presented phage displayed D (T)/S/(L) T K consensus sequence and contained peptides with pronounced glioma-selective properties. Phage clones expressing peptides with E (L)/V/(S) R G D S motif were found in cell lysates and represented the third family of glioma-specific ligands. All peptides within this family contain the RGD amino acid

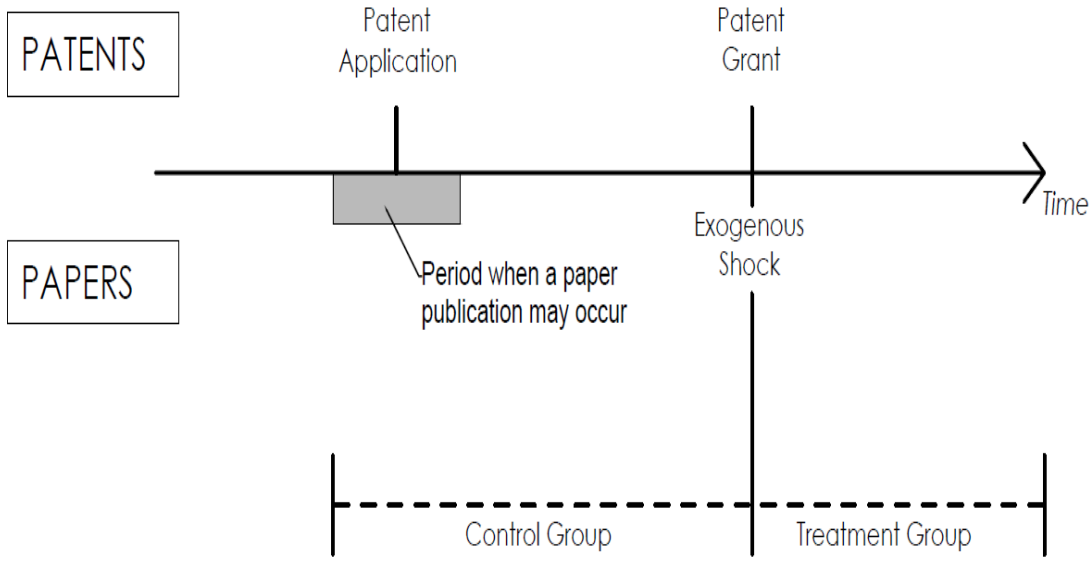
Source: apps.webofknowledge.com

between two populations is not applicable to our variables of interest, because of their skewed distribution. This aspect leads to the necessity of using an alternative framework that is simultaneously able to consider the non-normality of the entire set of variables and the heterogeneity shown by their distributions. For this reason, we opted for performing a *Mann-Whitney test* on median differences. This is a non-parametric test that compares the medians from two populations and it works for both continuous and discrete-count variables; as in our cases.

Second, we set up an econometric model able to isolate the effect of the patent granting on the rate of follow-on public knowledge. Here, the dataset is constituted by the 202 *patent-paper-pairs* resulting at the end of the identification process. As a proxy for the rate of follow-on public knowledge, we are going to use the citations that the paper receives after its publication by other peer-reviewed articles [6]. In particular, the main dependent variable is the amount of citations obtained by each publication on an annual base. The observation period goes from the paper publication year until the end

[6] This approach has been widely use in the literature. See, for example, de Solla Price (1965); Hall, Jaffe and Trajtenberg (2001); Posner (2000)

Figure 4.3: Scheme of the *patent-paper-pair* identification strategy



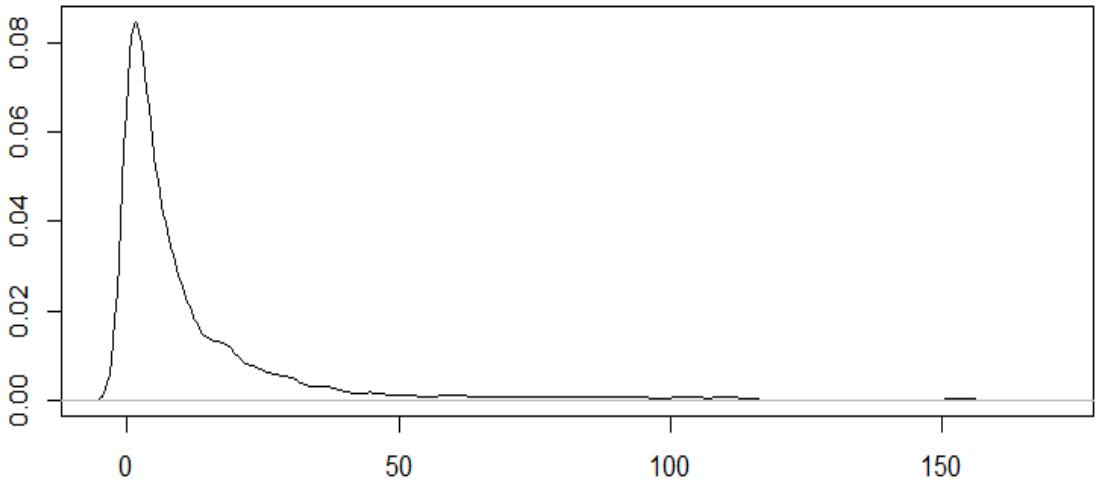
of 2014. The choice to truncate the count at this date has been done in order to have a complete information about each year. Since the 202 papers (obviously) have not been published all in the same year, we observe articles for periods with different lengths. Hence, the result is an unbalanced panel where the time of observation varies together with the age that every publication have reached at the end of 2014. In this structure, the final number of observations for the 202 publications is equal to 2536.

The biunivocal relation that we established between each patent and each paper contained in our dataset, allows for considering the grant of the patent as an *exogenous shock* in the paper's life. In fact, the granting occurs (on average) with a lag of 4.5 years after the application date. Since we construct the *pairs* requiring a correspondence between the patent's application date and the paper's publication period, roughly the same lag in time is shown also between the publication and the granting date. The presence of a shock, allows for the adoption of a *difference-in-difference* approach where the cita-

tions belonging to the pre-grant period represent the *control group* and the others form the *treatment group*. In order to formally distinguish the two sets, we are going to associate a binary variable to each observation assuming value equal to 1 if the year is in the post-grant period and 0 otherwise. Figure 4.3 supplies a graphical representation of the scheme we adopted to apply our model.

Recalling that our first aim is to identify the effect of a shifting in the IPR regime on the rate of follow-on public knowledge, we consider the annual number of citations received by each paper as dependent variable of our *difference-in-difference* model. Since this is a discrete counting variable highly right-skewed distributed (see Figure: 4.4), we have to consider the proper model fitting these characteristics and able to avoid heteroskedasticity.

Figure 4.4: Distribution of annual citations



Hausman, Hall and Griliches (1984) in their seminal paper state that there are two main regression models that may be used to deal with discrete count variable: (i) The Poisson regression model (PRM) and the (ii) The Negative Binomial regression model (NBRM). In order to choose the most fitting to our case, we need to compute the conditional mean and the conditional variance of the number of annual citations. In case the two are

not different, we can apply the Poisson regression method. However, our computations suggest that the mean and the variance assume respectively values of 12.24 and 410.89. The difference is clear and the over-dispersion is undeniable. This observation leads to the impossibility to apply the PRM to our case. Therefore, we opted for using the NBRM as the model able to better describe the characteristics of our dependent variable.

The estimation have been run using the *pglm()* command present in the *pglm library* available in the R-software.

4.3 Variables

In this section of the work we are going to describe all the variables that will be used in both the descriptive and regression analyses. In order to be clear as much as possible, we are going to gather them in four different sets, which are: (i) the dependent variable of the regression analysis (ii) Variables of interest (iii) paper variables and (iv) patent variables (See Table 4.1).

Dependent variable: Here we include the most important variable of the model represented by the number of annual citations ($ANNUAL\_CITS$) received by each one of the papers involved in a *pair*. Given the unbalanced structure of the panel, the total number of observations is 2536, that is the sum of the AGES of each publications. That is to say the number of years elapsed between the publication of the paper and 2014

Variables of interest: This set contains the two main explanatory variables of our model.
PAT_IN_FOR is a dummy which assumes value 0 in the years before the granting and value 1 in all the after-patent-grant years. This variable is the one that allows for dividing citations belonging to the *control group* from the ones filling up the *treatment-group*. Therefore, in the results' interpretation coefficient referring to this dummy deserves the highest level of importance.
$PATENT\_WINDOW$ is an other binary variable which assumes value equal to 1 only during the patent granting year. We retain it plays an important role for the explanation of the immediate effect caused by a shift in the IPR regime. These two variables can also be defined as pairs variables because

these are the only ones simultaneously referring to both patents and papers

Paper variables: Here we include a list of variables able to capture some publications' characteristics which will be used as control variables of our model.

PUB_ YEAR is the year during which the paper has been published in the peer-reviewed scientific journal.

N_ OF_ AUTHORS and *N_ OF_ ADDRESSES* respectively count the number of authors and the number of addresses appearing in a paper. While the first one gives information about the division of efforts behind an article, the second measures the number of different organizations involved in the production of the piece of knowledge disclosed through the paper.

PUB_ ADDR and *PRIV_ ADDR* are two binary variables that reflect the nature of the addresses' characteristics. Specifically, they assume value equal to 1 when at least one address is either public or private respectively. Among public institutions we consider universities, government organizations and laboratories[7].

IMP_ FACTOR is a measure for ranking, evaluating, categorizing, and comparing journals[8] yearly computed by ISI web of science. The impact factor is defined as the number of current year citations divided by the source items published in that journal during the previous two years. In this work we will use not the current, but the 2014 impact factor as a proxy for the quality of the journal. This choice is coherent with our decision to drop the observations at the end of the same year.

Patent-variables: Here we introduce a list of indicators available on the OECD-QUALITY DATABASE or calculated from PATSTAT DATABASE describing the patents' most important characteristics. Differently from the other variables described so far, these ones are calculated on the original set of 761 units rather than only on the dataset of PAIRED PATENTS used for our regression. It is worth noticing, that the utility of patent-variables is double: on one hand, we are going to use some of them for the comparison between

[7]The source we used to establish whether an address was public or not is the e-mail addresses, usually present in the paper.

[8]www.wokinfo.com

PAIRED and NO-PAIRED patents, on the other hand, these will be useful to add further specifications to our regression.

GRANT_YEAR: It is the year in which the USPTO recognises to the applicant the right of exclusivity for an invention granting a patent on it.

PATENT_SCOPE: It is associated with the technological and economic value of a patent. Lerner (1994) proposes a definition of patent scope based on the number of distinct 4-digit subclasses of the International Patent Classification (IPC). Each patent, indeed, is also classified according to an international classification reflecting different areas of technology[9]. Since each patent lists the international classes that have been assigned to, it become feasible to compute the *patent scope*.

$$PATENT\_SCOPE_p = n_p$$

where n_p is the number of IP classes listed in the patent. It has been argued that, the broader the scope index, the higher potential technological and market value of a patent[10].

FAMILY_SIZE: It represents the number of different patent offices where a given innovation has been registered

INV_TEAM_SIZE: The inventive time size indicates the number of patent's inventors. It is worth noticing that, for each patent, this is nothing but a subset of the papers' authors category. One of the three fundamental requirements for setting up a *patent-paper-pair* was, indeed, that all the inventors of the patents must appear as authors of the paper.

BWD_CITS: Backward citations refer to the list of patents, scientific works and other source of knowledge cited by a patent. Once an applicant has furnished this list, all the contained references are analysed by patent examiners who decide whether to keep or delete the citation. Criscuolo and Verspagen (2008) argue that the numbers of backwards citations is positively related with the novelty of a patent.

NPL_CITS: Citations of non-patent literature represent the list of refer-

[9]The IPC classification is available www.wipo.int

[10]Squicciarini, Dernis and Criscuolo (2013), p. 10

ences, different from patents, that have been cited. Differently from the backwards citations, the number of *NPL_CITS* has been interpreted as a good indicator not for novelty, but for the quality of the patent[11]. Moreover, the level of non patented literature citations reflects the contribution of public knowledge to a given invention [12].

NUM_OF_CLAIMS: The number and content of the claims determine the breadth of the rights conferred by a patent (OECD, 2009) and its expected value. In fact, patent fees are generally based on the amount of claims contained in the documents (the more the claims the higher the fees); therefore the patent's expected market value has to be related with the number of claims [13].

FWD_CITS5 and *FWD_CITS7* : Forward citations refer to the citations that a patent receives once being granted. Their number gives important information about the relevance that a patented invention has for the development of follow-on technologies. Since the measurement of forward citations can typically be done on two bases -i.e. five or seven years after the publication date- we describe two separate indicators for the two cases .

GENERALITY: The generality index relies on information concerning simultaneously forward citation and international classes (at every level of digit). It varies from 0 to 1 and it is computed as follows:

$$GENERALITY_X = 1 - \sum_{j=1}^{M_i} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{T_{ji}^n}{T_i^n} \right)^2$$

Where:

X= focal paper with $Y_{i=1,....N}$ citing papers

T_i^n = Total number of international classes (IPC) at the n-digit present in y_i

T_{ij}^n = Total number of IPC n-digit classes in the j^{th} IPC 4-digit classes in y_i

$j = 1...M_i$ is the cardinal of all IPC 4-digits classes in y_i

When the value of the index is close to 1, it means that the citing patents

[11] See Branstetter (2005)

[12] See Narin et al.(1997)

[13] See Tong and Davidson (1994)

belongs to a variety of different classes; meaning that the patent has a rather broad impact. On the contrary, a low level of $GENERALITY$ indicates that the patents that cites the x are concentrated in few fields of technologies.
$ORIGINALITY$: Patent originality refers to the breadth of the technology fields on which a patent relies[14]. The logic to build the indicator is quite similar to the one used to set up the generality index. However, now the focus in on the backward- rather than the forward- citations. Analytically, the index can be computed as follows:

$$ORIGINALITY_X = 1 - \sum_{j}^{n_x} s_{pj^2}$$

Where s_{pj} is the percentage of citations made by patent$X$ to technological class j out of the the total n_x patent classes. The rationality of this indicator is that the wider is the knowledge, the larger is the potential for originality. As in the generality index case, also the originality index is defined between 0 and 1.

Table 4.1 furnishes a scheme useful to clarify the variables distinction and definitions.

[14]Squicciarini, Dernis and Criscuolo (2013), pag. 49

Table 4.1: Variables description

Variables	Description	Source
Dependent variable		
ANNUAL_CITS	Count of citations yearly recived by the focal paper after the publication date	ISI
Pairs variables		
PAT_IN_FOR	Dummy assuming value 1 if the citation is recived in the post-grant period	USPTO
PATENT_WINDOW	Dummy assuming value 1 if the citation is recived in the year of grant	USPTO
Papers variables		
PUB_YEAR	Year of publication in the journal	ISI
N_OF_AUTHORS	Number of authors of the article	ISI
N_OF_ADDRESSES	Number of institutions involved in the publication	ISI
PUB_ADDR	Dummy set to 1 if at least 1 address is public	ISI
PRIV_ADDR	Dummy set to 1 if at least 1 address is private	ISI
IMP_FACTOR	Imact factor of the journal	ISI
Patents variables		
GRANT_YEAR:	Year in which the patent is granted	USPTO
PATENT_SCOPE	Number of international classes in the patent	OECD
FAMILY_SIZE:	Number of patent offices	OECD
INV_TEAM_SIZE	Number of inventors	PATSTAT
BWD_CITS	Number of citations made by a patent	OECD
NPL_CITS	Citations of non patented literature	OECD
N_OF_CLAIMS	Number of listed claims	OECD
FWD_CITS5	Citations recived in 5 years after the grant	PATSTAT
FWD_CITS7	Citations recived in 7 years after the grant	OECD
GENERALITY	Index reflecting generality of a patent	OECD
ORIGINALITY	Index reflecting originality of a patent	OECD

Chapter 5

Results

5.1 Descriptive analysis

We begin our analysis investigating the main statistical properties of the entire list of variables described in the previous section. It is important to note from the beginning that the number of observations varies across variables, depending on their nature. Specifically, *ANNUAL_CITS*, *PAT_IN_FOR* and *PATENT_WINDOW* have been annually computed for each *patent-paper-pair*, hence their cardinality corresponds to the rows of our unbalanced panel (2536). **Paper variables**, instead, are constant during the years for each publication, therefore we have a set 202 observations for each variable. Lastly, **Patent variables** can be computed both for PAIRED and NO-PAIRED patents reaching a total number of 761 observations ($202 + 559$).

Variables of interest statistics

Table 5.1: Variables of interest statistics

Statistic	N	Mean	St. Dev.	Min	Max
ANNUAL_CITS	2536	12.245	20.271	0	166
PAT_IN_FOR	2536	0.613	0.487	0	1
PATENT_WINDOW	2536	0.080	0.271	0	1

Table 5.1 shows the main statistical properties of the *ANNUAL_CITS*, *PAT_IN_FOR* and *PATENT_WINDOW*. The fact that the mean value of *PAT_IN_FOR* is equal to 0.613 suggests that we are dealing with publications that have been associated (on average) with a patent for a period slightly longer than the age they have reached at the end of 2014. Simultaneously, the very low mean value of *PATENT_WINDOW* (0.080) is simply explained by the way we used to build this variable since we assume that it takes value one only if the citation is made during the granting-patent year. Lastly, the mean value of *ANNUAL_CITS* (12.245) is particularly meaningful when it is compared with the median (5) and with the variance (410.9134). The distance that occurs between the mean and the variance tells us something about the shape of the variable's distribution, which, in this case, is highly right-skewed. The comparison between the the mean value and the variance, instead, is essential in order to understand the level of dispersion, that is particularly high in our instance.

Paper variables statistics

The main statistical properties about publications characteristics are summarised in Table 5.2. The fist observation deserving a note regards the boundary values assumed by the *PUB_YEAR*. The oldest publication we included in our dataset of PAIRED patents dates back to 1986. Since we begun our screening process for the identification of the pairs including patents non granted before 2004, it may appear as a contradiction to the *date-matching-requirement*. However, when considering the possibility of a match not only with the application-date, but also with a previous filling-date, the result should not come as a surprise. On the other hand, the most recent publication year is 2010. In our scheme, it indicates a distance between the application and the granting date very short. This situation is referred to those patents which are recognised as highly qualitative and deserving a priority review in order to reduce the period that usually occurs before the approval. However, extreme cases are rare in our set as we observe just 5 and 2 cases (out of 202) of papers published respectively before 1998 and

after 2009.

Table 5.2: Paper variables statistics

Statistic	N	Mean	St. Dev.	Min	Max
PUB_YEAR	2536	2002.5	3.639	1986	2010
N_OF_AUTHORS	202	9.14	5.209	2	31
N_OF_ADDRESSES	202	4.01	2.514	1	18
PUB_ADDR	202	0.94	0.237	0	1
PRIV_ADDR	202	0.28	0.454	0	1
IMP_FACT	202	10.10	11.322	1.380	55.873

The *N_OF_ADDRESSES* is, on average, 4. The statistics referring to their division in *PUB_ADDR* and *PRIV_ADDR* contribute to give a clear idea about the nature of the organizations involved in the publications. Recalling that the private or public characterization has been given after the check that at lest one address is either private or public, the high mean value of the dummy *PUB_ADDRS* (0.94) suggests that the almost the totality of papers composing our set has a link with either a university or an other public institution. This confirm the importance of public institutions in the development of knowledge in this field. On the other hand the percentage of private organizations involved in the publication path is, on average, very low (28%).

Lastly, observing the properties of *IMP_FACT*, we found that the articles involved in our analysis belong to a set of journals very heterogeneous in terms of quality. This result is supported by both the high value of the standard deviation (11.322) and by the long distance between the max and the min values. Moreover, we can go a little further comparing the mean value of *IMP_FACT* (10.1) with the mean value of annual citations received by the papers (12.245. See Table 5.1). Recalling that the impact factor is computed by the ISI as an average of the citations yearly received by the articles contained in a specific journal, the difference of 2.23 may be interpreted has the sign that the publications included in our set of *pairs*, are (on average) annually more cited than articles collected in the same journal.

Patent variables statistics

In Table 5.3 we report a summary of the main statistical properties of patent variables for the entire initial set of 761 patents, regardless of their division in PAIRED and NO-PAIRED.

Table 5.3: Patent variables statistics

Statistic	N	Mean	St. Dev.	Min	Max
GRANT_YEAR	761	2008	2.305	2004	2011
PATENT_SCOPE	761	3.524	1.949	1	11
FAMILY_SIZE	761	7.505	6.138	1	38
INV_TEAM_SIZE	761	3.152	1.896	1	11
BWD_CITS	761	15.763	22.447	0	113
NPL_CITS	761	33.311	30.155	0	125
N_OF_CLAIMS	761	15.538	15.914	1	190
FWD_CITS5	761	3.136	5.799	0	89
FWD_CITS7	761	4.201	7.326	0	110
GENERALITY	761	0.641	0.218	0.000	0.924
ORIGINALITY	761	0.845	0.144	0.000	0.963

Although it may be interesting to in depth discuss these characteristics in order to have an idea about the peculiarities of the patents contained in the technological class 435/7.23, this is not the aim of our work. Rather, we aim at identifying differences between PAIRED and NO-PAIRED patents included in the class. Therefore, the next section achieves at underlying those differences at a descriptive level.

5.1.1 A comparison between PAIRED and NO-PAIRED patents

The variables that we intend to consider for the sake of comparison are summarised in Tables 5.4 and 5.5, which refer to the sets of NO-PAIRED and PAIRED patents respectively.

In order to test the differences between the properties of the two sets, we have to take into consideration that the variables' distributions are not normal. As an example, Figures 5.1 and 5.2 show the density function of

Table 5.4: Characteristics of NO-PAIRED patents

Statistic	N	Mean	St. Dev.	Min	Max
PATENT_SCOPE	559	3.605	2.005	1	11
BWD_CITS	559	16.446	22.444	0	113
NPL_CITS	559	32.832	30.224	0	125
N_OF_CLAIMS	559	15.570	15.309	1	148
FWD_CITS5	559	3.245	6.094	0	89
FWD_CITS7	559	4.318	7.588	0	110
GENERALITY	559	0.645	0.218	0.000	0.924
ORIGINALITY	559	0.848	0.143	0.000	0.963

Table 5.5: Characteristics of PAIRED patents

Statistic	N	Mean	St. Dev.	Min	Max
PATENT_SCOPE	202	3.302	1.771	1	8
BWD_CITS	202	13.891	22.406	0	104
NPL_CITS	202	34.624	30.001	0	110
N_OF_CLAIMS	202	15.450	17.506	1	190
FWD_CITS5	202	2.837	4.903	0	30
FWD_CITS7	202	3.881	6.561	0	39
GENERALITY	202	0.627	0.219	0.000	0.873
ORIGINALITY	202	0.836	0.149	0.000	0.958

NPL_CITS and of *ORIGINALITY* both in cases of NO-PAIRED and PAIRED patents.

Figure 5.1: *NPL_CITS* density for PAIRED(red) and NO-PAIRED(black) patents

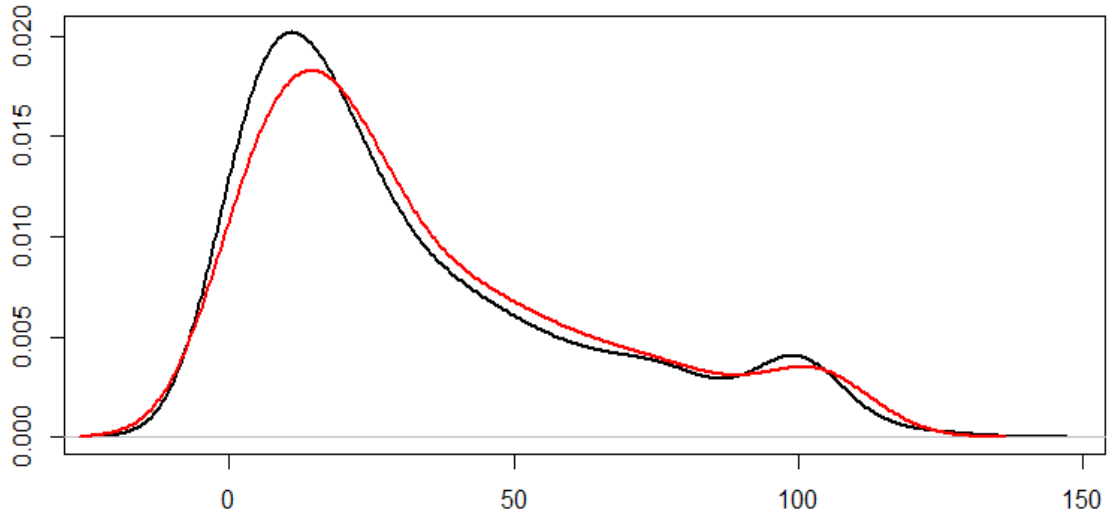
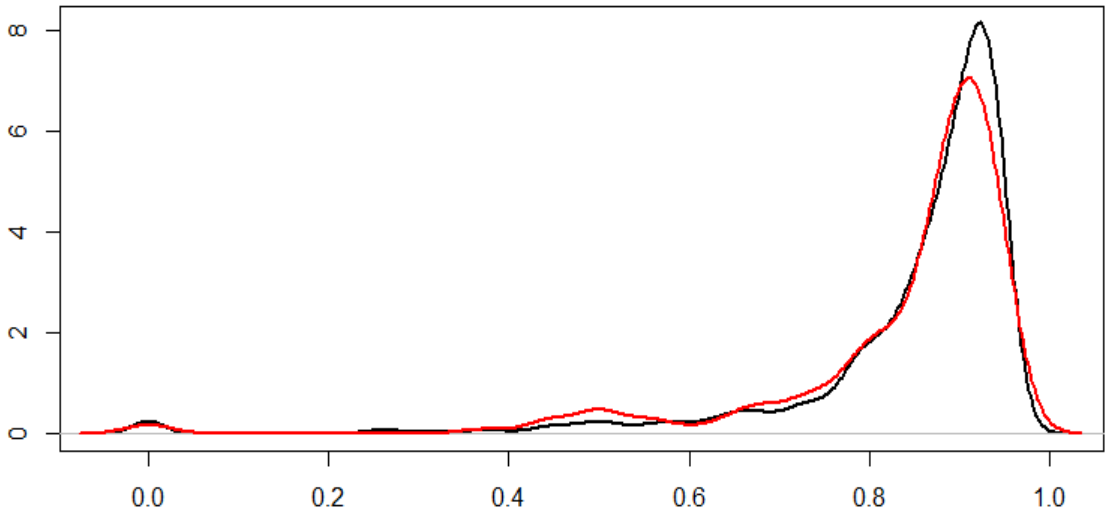


Figure 5.2: *ORIGINALITY* density for PAIRED(red) and NO-PAIRED(black) patents



Although the shapes of the densities are very different, it is undoubtedly

true that in the two cases the distribution is not-normal. The non patented citations clearly present a right-skewed distribution, the originality index, instead, has a left-skewed distribution. The observed non regularity and non normality impedes to use a classical *t-test* on mean differences in order to compare the two sets of variables.

Therefore, we opted to perform a *Mann-Whitney test* on median differences. In Table 5.6 we report the medians, for each indicators, from both the NO-PAIRED and PAIRED patents sets. Moreover, in the third column, we indicate the p-value associated to the performed *Mann-Whitney test*. In particular, the null hypothesis we want to test is that there is no difference between the medians from the two sets.

Table 5.6: Mann-Whitney test on median difference

Statistic	Median NPP	Median PP	M-W test
PATENT_SCOPE	3	3	(0.1209)
BWD_CITS	8	5	(0.0004)***
NPL_CITS	22	25	(0.0274)*
N_OF_CLAIMS	11	12	(0.9317)
FWD_CITS5	1	1	(0.9317)
FWD_CITS7	2	1	(0.0417)*
GENERALITY	0.714323	0.7044291	(0.2349)
ORIGINALITY	0.8960	0.8292	(0.0163)**

Note: *p<0.1; **p<0.05; ***p<0.01

The variables which present a significant difference between the two sets are *BWD_CITS, NPL_CITS, FWD_CITS7* and *ORIGINALITY* . On the other hand, *PATENT_SCOPE, N_OF_CLAIMS, FWD_CITS5* and *GENERALITY* show not statistically different medians in the NO-PAIRED PATENTS and PAIRED PATENTS cases.

The results of the *NPL_CITS* is not surprising because of the sample construction. Both the median and the mean are higher in the PAIRED PATENTS (32.8 vs. 34.6 and 22 vs.25). As already specified, *NPL_CITS* are a good indicator for the contribution of public knowledge to a specific innovation. Given the criteria that we used to establish a *patent-paper-pair*,

it becomes evident that the inventors of a paired patents are also involved in the *academic world*. Therefore, the fact that they have an inclination to cite public articles should not come as a surprise. However, Branstetter (2005) advocates that *NPL_CITS* are also a measurement positively related with the quality of the patent itself.

Among the various indicators we considered, important ones for the sake of quality are *BWD_CITS, FWD_CITS5, FWD_CITS7, ORIGINALITY* and *GENERALITY*. From the analysis of the results we got, it seems that their values tends to contradict the Branstetter's hypothesis, when the (very specific and limited) *patent-papers-pair* analysis is carried out. In fact, mean and median of these three indicators follow a path which goes it the opposite direction. The highest values, indeed, are observed in case of NO-PAIRED patents. In particular, the *BWD_CITS* show a difference at the highest level of significance in favour of the NO-PAIRED PATENTS. Since it is a wide-shared opinion[1] that *BWD_CITS* are positively correlated with the degree of novelty of the invention, we shall state that the PAIRED patents in our sample appear to be less novel than the NO-PAIRED ones.

When we combine information about *BWD_CITS* and technological classes (IPC) together, we found out that PAIRED patents also sin in terms of *ORIGINALITY*. Although the values of the originality index are quite high in both cases, there is a significant evidence that NO-PAIRED patents are more original than the PAIRED ones. Following the common interpretation, it means that the latter group of observations cites document classified in a smaller number of technological classes.

Lastly, the analysis of differences in the forward citations, suggests that, when the basis of observation is 7 years (*FWD_CITS7*), PAIRED patents, plausibly assumes a lower relevance for the development of follow-on-technology, when a comparison with the NO-PAIRED PATENTS is done.

The totally different path observed for *NPL_CITS*, on one side, and for *BWD_CITS*, *FWD_CITS7* and *ORIGINALITY*, on the other, seems to lead to an inconsistency. However this incongruity is very likely to be restricted at the specific case of study. In particular, the *patent-paper-pair*

[1] Verspagen (2008)

approach itself may be the cause of it. In fact, when it is recognised that a patent is associated with a previously published paper, the piece of knowledge disclosed after the granting is actually already in the public domain through publications. This may be a good explanation for the reason why we observe, in our sets, lower level of *FWD_CITS, BWD_CITS* and *ORIGINALITY* in case of PAIRED PATENTS. On the other hand, the high value of *NPL_CITS* is consistent with this kind of analysis since, regardless of the period, publishers in scientific journals keep to be willing to cite academic works even when they *becomes inventors.*

Despite the potential importance of the result, it is worth noticing that the analysis of the patent characteristics has been conducted at a descriptive level. It would be totally misleading to generalise this conclusion at a wider dimension, but it constitutes a good starting point for future research.

5.2 Econometric Analysis

The parameters we are going to estimate by means of NBRM are showed in Equation 5.1

$$ANNUAL\_CITS_{i,t} = f(\alpha PATENT\_WINDOW_{i,t} + \beta PAT\_IN\_FOR_{i,t} + \gamma N\_OF\_AUTHORS_i + \delta N\_OF\_ADDRESSES_i + \eta PUB\_ADDRESS_i + \zeta PRIV\_ADDRESS_i + \psi IMP\_FACT_i; \epsilon_{i,t}) \tag{5.1}$$

In particular, we focus on α and β, which are the coefficients of our explanatory variables. The results of this first regression are reported in Table 5.7. Here, we observe that, among the control variables, *PRIV_ADDR* and *N_OF_ADDRESSES* are significantly correlated with the number of annual citations. Specifically, an higher number of addresses lead to an increase in the rate of follow-on publications. On the other hand, when the nature of addresses is private, it implies a reduction it the number of *ANNUAL_CITS*. The latter observation is an interesting one. In fact, it suggests that even in

case of public tools of disclosure (papers), the private nature of the organizations involved in the publication path can reveal to be slightly damaging for future research.

Going back to our variables of interest, we observe that *PAT_IN_FOR* does not significantly influence the *ANNUAL_CITS* count. On contrary *PATENT_WINDOW* positively affect the number of citations yearly received by the focal paper[2]. Both the results are in contrast with our original expectation based on Huang's and Murray's (2009) seminal paper.

In that paper, the authors apply the *patent-paper-pair* approach in case of research on human genetics. Although the set of hypotheses they aim to test is much broader than ours, the first assumption that is verified in the paper is that patent grants have a strong (17%) negative impact for the rate of annual citations. The difference between the results could be driven by several factors. In fact, it is worth noticing that our work is only partially comparable with Huang's and Murray's work. First of all, the *abstract-matching-requirement* is much more stringent and unmistakable when it is conducted in the genetic fields. The starting set considered by the authors, indeed, is not a specific technological class, but all the patents disclosing a gene sequence. The latter must be also the novelty diffused in the paper. It is clear that when the innovation is directly referred to a *codified information* (i.e. the gene sequence), it becomes much more identifiable as the possible related infringements do. However, a part from this observation, we tried to give a further interpretation to the obtained result which will be extensively discussed in the next section.

5.2.1 A possible explanation

In order to understand the nature of our opposite results, we further investigate the nature of the forward citations received by the 202 paired-publications.

Specifically, we focused on the identities of the citing papers' authors. At

[2]Specifically, the effect is an increase of 31%. The latte is obtained exponentiating the coefficient in Table 5.7 and subtracting 1 from the result

Table 5.7: Results equation 5.1

	Dependent variable:
	ANNUAL_CITS
PATENT_WINDOW	0.274997
	(9.12e-09)***
PAT_IN_FOR	-0.009126
	(0.766630)
N_OF_AUTHORS	-0.007829
	(0.446488)
N_OF_ADDRESSES	0.053334
	(0.020629)*
PRIV_ADDRESS	-0.131465
	(0.0257680)*
PUB_ADDRESS	0.001065
	(0.766664)
IMP_FACT	0.155631
	(0.443905)
Observations	2536
Log-likellihood	-6240.318
Note:	*p<0.1; **p<0.05; ***p<0.01

first glance, we observed that a considerable share of citations had been made by the patents' inventors. Given the hight relevance that this feature will assume for this work, we define a new variable that has never been considered earlier in the literature. We refer to auto-citations ($AUTO\_CITS$) as those ones that, when a *patent-paper-pair* is considered, the paper receives by the paired patent's author.

The identification process for the $AUTO\_CITS$ is neither automatic nor immediate. In fact, there are no alternative ways for the analysis of the citation path associated to each publication, if not manually matching data from the USPTO and the ISI WEB OF SCIENCE. Our aim is to understand how the auto citations behave before and after the granting in a way which is formal as much as possible. However, to be coherent with the analytical framework that we used so far, we should be able to identify not only the total number of auto citations received by each paper, but also their *annual distribution*. Since we are facing a total amount of citations equal to 31053 observations, it would be hard, if not impossible, to manually classify them for authors and publication year.

Our expectation is that the the portion of $AUTO\_CITS$ is significantly higher after the grant, since the patent's inventors would be the only ones not risking to incur in legal suits citing the paper contained a piece of knowledge covered by IPR. In order to test this hypothesis, being unable to consider the entire set of observations for the reasons explained above, we consider a limited subset of publications. In particular, we used the R-command *sample(20)* to randomly choose twenty *patent-paper-pairs* whose citations have been deeply analysed. We found out that in 17 cases out of 20, the share of annual citations post-grant is higher than the same measurement in the pre-granting period. Specifically, it emerges that, on average:

- The percentage of auto citations in the first period is 5%
- The percentage of auto citations in the second period is 20%

Despite the small size of the sample, the trend is clear and unambiguous. Therefore, we include this feature in our model by approximating the difference in the number of $AUTO\_CITS$ pre and post grant with a value of 15%.

Moreover, given the significant difference inside the sample, we extended this result to the entire set of *patent-paper-pairs* and we built an other variable (NET_CITS) which will be our new dependant variable and is defined as follows:

$$NET\_CITS_{i,t} = \begin{cases} (1-\bar{p})ANNUAL\_CITS_{i,t} & \text{if } PAT\_IN\_FOR_{i,t} = 0 \\ (1-\bar{P})ANNUAL\_CITS_{i,t} & \text{if } PAT\_IN\_FOR_{i,t} = 1 \end{cases}$$

Where:
$\bar{p}$ is a constant approximating the mean value of the pre-grant $AUTO\_CITS$
$\bar{P}$ is a constant approximating the mean value of the post-grant $AUTO\_CITS$

In our case, $\bar{p}$ and $\bar{P}$ assume values 0.05 and 0.15 respectively. Therefore, by applying the definition of NET_CITS, we proceed to re-estimate model 5.1 considering the estimated number of net citations (instead of $ANNUAL\_CITS$) as dependent variable (See equation 5.2).

$$\begin{aligned} NET\_CITS_{i,t} = f(\alpha PATENT\_WINDOW_{i,t} + \beta PAT\_IN\_FOR_{i,t} \\ +\gamma N\_OF\_AUTHORS_i + \delta N\_OF\_ADDRESSES_i + \eta PUB\_ADDRESS_i \\ +\zeta PRIV\_ADDRESS_i + \psi IMP\_FACT_i; \epsilon_{i,t}) \end{aligned} \tag{5.2}$$

The estimated coefficients for our two explanatory variables are shown in Table 5.8. Here, in the post-grant period the expected net annual rate of forward citations is negative at the highest level of significance with a decline of 11.67 %[3]. However, the $PATENT\_WINDOW$ coefficient keep being positive indicating an increase of net citations during the granting year.

Although our original work did not aim to consider the separate effect

[3]In order to interpret the coefficients as the percent change of the dependant variable corresponding to a unit change of the dependant variable, it has to be exponentiated the factor directly obtained from the negative binomial regression and, then, it has to be deducted 1. Therefore, the explicit calculation we performed to get the measurement of the decline are $exp(-0.12414) - 1 = -0.1167$

of granting on the patent's owners and on third parties, the first wave of (unexpected)[4] results gave us the right spark to deeper investigate which are the implications of IPR regimes in the field of medical research. What emerges from our analysis is that, even though the total amount of follow-on public knowledge (measured by means of forward $ANNUAL\_CITS$) does not decline after the grant, it has an important effects on third parties. In fact, their attitude to cite the focal paper is relatively lower in the second period when they incur the risk of legal suits. On the contrary, the patent's authors, are the only ones free to keep citing the article with no-infringement-risks. Therefore, we argue that the effect of granting is on the concentration of knowledge in the hands of few actors, rather than on its absolute quantity.

Table 5.8: Results equation 5.2

	Dependent variable:
	NET_CITS
$PATENT\_WINDOW$	0.26342
	(6.69e-08)***
PAT_IN_FOR	-0.12414
	(4.62e-05)***
Observations	2536
Log-likellihood	-6067.097
Note:	*p<0.1; **p<0.05; ***p<0.01

Furthermore, we observe that the decline in the rate of annual citations does not occur in the patent granting year. It may be well explained if we consider that the grant comes to third parties as a surprise. Therefore, the impact of a shift in the institutional environment cannot be immediately evident. On the contrary, this lag strengthens our hypothesis of *knowledge concentration*. The fact that in the following years the drop of net citations

[4]Being our work based on the *patent-paper-pair* approach, we initially expected to find results coherent with other studies conducted using the same methodology, i.e. a decline in the number $ANNUAL\_CITS$ associated with the practice of granting.

clearly manifests, might be the real consequence for legal costs ascribed to third parties that have cited the focal paper during the patent grant year. Obviously, this is just an hypothesis that should be verified using data referring to infringement costs.

The last step of our econometric work, consists of including in the model some patents characteristics in order to underline their interaction with the effect of PAT_IN_FOR on the NET_CITS annual trend. In particular, we analysed the nature and the intensity of the possible impacts that $PATENT\_SCOPE$, $FAMILY\_SIZE$, N_OF_CLAIMS and INV_TIME_SIZE separately have on the measurement of the decline in NET_CITS during the after-grant years. However, the first three variables seem not to play any significant role on this matter[5].

Table 5.9: Model with INV_TEAM_SIZE

	Dependent variable:
	NET_CITS
$PATENT\_WINDOW$	0.26241
	(7.67e-08)***
PAT_IN_FOR	-0.15435
	(4.44e-05)***
INV_TEAM_SIZE	0.01917
	(0.535801)
Observations	2536
Log-likellihood	-6066.905
Note:	*p<0.1; **p<0.05; ***p<0.01

On the contrary, we found an interesting result when we include the inventive team size. In this case, the decline of net forward citation in the second period amounts to 14.3% (See Table 5.9), which means an extra

[5]The variation of the coefficient of PAT_IN_FOR never exceeds the |0.01%| when one these variable is added to our general model (equation 5.2)

drop of 2.6%. It is worth noticing that this result is a further validation of our *concentration hypothesis*. The inventive team size, indeed, expresses the number of patent's inventors; the higher it is, the higher the probability of observing a reduction in the rate of net forward citations becomes.

Conclusions

The aim of this work was to investigate the effects that the practice of granting patents at the highest level of medical research has on the follow-on *public knowledge*. In fact, the growing number of patents referring to genetic knowledge contributes to fuel the debate over the role of IPR regimes as a policy either able to spur innovation or responsible to damage technical progress. A number of scholars (Orsi and Coriat, 2005) have considered single case studies showing some evidences about the negative effects of gene patenting for the long run supply of *public knowledge*. The Myriad genetics case, ended with the invalidation of two genes referring to ovarian cancer, is exemplar in this sense.

In order to expand the same kind of analysis at a broader level- not focusing on specific cases- we adopted a *patent-paper-pair* identification approach. Thanks to it, we were able to concretely identify pieces of knowledge disclosed by both private and public tools of disclosure. In the specific case of research on cancer, we found that using the *hybrid* instrument of *patent-paper-pair* as a tool of disclosure, is far from being an rare practice. In fact, 27% of patents of our original set are undoubtedly associated with a paper.

Previous studies applying the same methodology are mostly due to Murray's work (See Murray (2002) and Huang and Murray (2009)) that we used as a benchmark for our analysis. However, differently from the author, we chose to study the IPRs implications when a specific therapeutic area (research on cancer) is taken into consideration.

Contrary to our original expectation we found out that the effect of patent granting is on the different types of citing authors, rather than on the amount of follow-on *public knowledge*. In other words, granting a patent on a piece

of knowledge already available in the public domain does not have an impact on the total number of forward citations received by the paper. However, patents' inventors and third parties show different citing behaviours before and after the grant. In fact, considering the number of net citations (i.e. those ones received by third parties) we registered an expected decline in citations equal to 11.67% after the patent granting.

Although this result has been obtained considering only a subset of patents included in our sample, it finds a convincing validation when other two aspects are considered. Firstly, we observed that the decline in net citations is positively correlated with the number of patent's inventors. Secondly, the drop in the rate of net citations is not observed during the grant year, but clearly manifests in the following period. It may reasonably occur because the patent grant comes as a surprise for third parties who would keep citing the paper without the awareness of the infringement. Therefore, we claim that an analysis of the infringement costs ascribed to third parties, is a good starting point for future research agenda.

Moreover, our findings contribute to reveal some important differences between paired and unpaired patents. Specifically, we found that, when intellectual property rights are granted on pieces of knowledge already publicly disclosed, the resulting patents appear to be less original and less cited than the other (unpaired) patents contained in the same technological class. On the contrary, we observed that papers considered in our sample of pairs, are, on average, more cited than other papers published by the same journal. We claim that a further investigation about this controversial evidence is necessary for a deeper comprehension of the role played by private tools of disclosure in case of basic medical research.

Bibliography

[1] Marcia Angell. Excess in the pharmaceutical industry. *Canadian Medical Association Journal*, 171(12):1451–1453, 2004.

[2] Ashish Arora and Alfonso Gambardella. The changing technology of technological change: general and abstract knowledge and the division of innovative labour. *Research policy*, 23(5):523–532, 1994.

[3] Kenneth J Arrow. The economic implications of learning by doing. *The review of economic studies*, pages 155–173, 1962.

[4] James Bessen and Eric Maskin. Sequential innovation, patents, and imitation. *The RAND Journal of Economics*, 40(4):611–635, 2009.

[5] James Bessen and Michael James Meurer. *Patent failure: How judges, bureaucrats, and lawyers put innovators at risk*. Princeton University Press, 2008.

[6] James E Bessen. Patent thickets: Strategic patenting of complex technologies. *Available at SSRN 327760*, 2003.

[7] Michele Boldrin, D Levine, and Alessandro Nuvolari. Do patents encourage or hinder innovation? the case of the steam engine, 2008.

[8] Michele Boldrin, David K Levine, et al. Against intellectual monopoly. 2008.

[9] Lee Branstetter. Exploring the link between academic science and industrial innovation. *Annales d'Economie et de Statistique*, pages 119–142, 2005.

[10] Vannevar Bush. Science: The endless frontier. *Transactions of the Kansas Academy of Science (1903)*, pages 231–264, 1945.

[11] Kenneth Carow, Randall Heron, and Todd Saxton. Do early birds get the returns? an empirical investigation of early-mover advantages in acquisitions. *Strategic Management Journal*, 25(6):563–585, 2004.

[12] Mildred K Cho, Samantha Illangasekare, Meredith A Weaver, Debra GB Leonard, and Jon F Merz. Effects of patents and licenses on the provision of clinical genetic testing services. *The Journal of molecular diagnostics*, 5(1):3–8, 2003.

[13] Mario Cimoli, Giovanni Dosi, Keith E Maskus, Ruth L Okediji, Joseph E Stiglitz, and Jerome H Reichman. *Intellectual property rights: legal and economic challenges for development*. Oxford University Press, 2014.

[14] Mario Cimoli, Giovanni Dosi, and Joseph E Stiglitz. The political economy of capabilities accumulation: The past and future of policies for industrial development. *Mario Cimoli, Giovanni Dosi and Joseph E. Stiglitz*, 2009.

[15] Public Citizen. Rx r&d myths: The case against the drug industry's r&d 'scare card', 2001.

[16] Wesley M Cohen, Richard R Nelson, and John P Walsh. Protecting their intellectual assets: Appropriability conditions and why us manufacturing firms patent (or not). Technical report, National Bureau of Economic Research, 2000.

[17] Benjamin Coriat, Fabienne Orsi, and Cristina d'Almeida. Trips and the international public health controversies: issues and challenges. *Industrial and corporate change*, 15(6):1033–1062, 2006.

[18] Paola Criscuolo and Bart Verspagen. Does it matter where patent citations come from? inventor vs. examiner citations in european patents. *Research Policy*, 37(10):1892–1908, 2008.

[19] Edward L Deci, Richard Koestner, and Richard M Ryan. Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of educational research*, 71(1):1–27, 2001.

[20] Vishwas Devaiah. A history of patent law. *Retrieved August*, 10:2010, 1992.

[21] Joseph A DiMasi, Ronald W Hansen, and Henry G Grabowski. The price of innovation: new estimates of drug development costs. *Journal of health economics*, 22(2):151–185, 2003.

[22] Giovanni Dosi. Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research policy*, 11(3):147–162, 1982.

[23] Giovanni Dosi, Luigi Marengo, and Corrado Pasquali. How much should society fuel the greed of innovators?: On the relations between appropriability, opportunities and rates of innovation. *Research Policy*, 35(8):1110–1121, 2006.

[24] Giovanni Dosi, Orietta Marsili, Luigi Orsenigo, and Roberta Salvatore. Learning, market selection and the evolution of industrial structures. *Small Business Economics*, 7(6):411–436, 1995.

[25] Giovanni Dosi and Richard R Nelson. Technical change and industrial dynamics as evolutionary processes. *Handbook of the Economics of Innovation*, 1:51–127, 2010.

[26] Giovanni Dosi, Richard R Nelson, and Sidney G Winter. Introduction: The nature and dynamics of organizational capabilities. *The nature and dynamics of organizational capabilities*, 1:24, 2000.

[27] Daniel C Fehder, Fiona Murray, and Scott Stern. Intellectual property rights and the evolution of scientific journals as knowledge platforms. *International Journal of Industrial Organization*, 36:83–94, 2014.

[28] Christopher Freeman and Luc Soete. *The economics of industrial innovation*. Psychology Press, 1997.

[29] Jeanne C Fromer. Patent disclosure. *Iowa L. Rev.*, 94:539, 2008.

[30] Nancy T Gallini and Ralph A Winter. Licensing in the theory of innovation. *The RAND Journal of Economics*, pages 237–252, 1985.

[31] Alfonso Gambardella. *Science and innovation: The US pharmaceutical industry during the 1980s*. Cambridge University Press, 1995.

[32] Ove Granstrand. Internationalization of corporate r&d: a study of japanese and swedish corporations. *Research Policy*, 28(2):275–302, 1999.

[33] Debra L Greenfield. Greenberg v. miami children's hospital: Unjust enrichment and the patenting of human genetic material. *Annals Health L.*, 15:213, 2006.

[34] Bronwyn H Hall, Adam Jaffe, and Manuel Trajtenberg. Market value and patent citations. *RAND Journal of economics*, pages 16–38, 2005.

[35] Bronwyn H Hall, Adam B Jaffe, and Manuel Trajtenberg. The nber patent citation data file: Lessons, insights and methodological tools. Technical report, National Bureau of Economic Research, 2001.

[36] Bronwyn H Hall and Rosemarie Ham Ziedonis. The patent paradox revisited: an empirical study of patenting in the us semiconductor industry, 1979-1995. *RAND Journal of Economics*, pages 101–128, 2001.

[37] Garrett Hardin. The tragedy of the commons. *science*, 162(3859):1243–1248, 1968.

[38] Dietmar Harhoff, Frederic M Scherer, and Katrin Vopel. Citations, family size, opposition and the value of patent rights. *Research Policy*, 32(8):1343–1363, 2003.

[39] Jerry A Hausman, Bronwyn H Hall, and Zvi Griliches. Econometric models for count data with an application to the patents-r&d relationship, 1984.

[40] Michael Heller. The gridlock economy. *sesquicentennial essays of the faculty of columbia law school*, page 127, 2008.

[41] Michael A Heller and Rebecca S Eisenberg. Can patents deter innovation? the anticommons in biomedical research. *Science*, 280(5364):698–701, 1998.

[42] Kenneth G Huang and Fiona E Murray. Does patent strategy shape the long-run supply of public knowledge? evidence from human genetics. *Academy of Management Journal*, 52(6):1193–1221, 2009.

[43] Pia Hurmelinna-Laukkanen and Kaisu Puumalainen. Nature and dynamics of appropriability: strategies for appropriating returns on innovation. *R&d Management*, 37(2):95–112, 2007.

[44] Adam B Jaffe. The us patent system in transition: policy innovation and the innovation process. *Research policy*, 29(4):531–557, 2000.

[45] Adam B Jaffe and Josh Lerner. Patent prescription: a radical cure for the ailing [us patent policy]. *Spectrum, IEEE*, 41(12):38–43, 2004.

[46] Sunil Kanwar and Robert Evenson. Does intellectual property protection spur technological change? *Oxford Economic Papers*, pages 235–264, 2003.

[47] Samuel Kortum and Josh Lerner. Stronger protection or technological revolution: what is behind the recent surge in patenting? In *Carnegie-Rochester Conference Series on Public Policy*, volume 48, pages 247–304. Elsevier, 1998.

[48] Thomas S Kuhn. The structure of scientific revolutions vol. 1962.

[49] John Langrish. *Wealth from knowledge: a study of innovation in industry*. Halsted Press, 1972.

[50] Josh Lerner. 150 years of patent protection. Technical report, National bureau of economic research, 2000.

[51] Joshua Lerner. The importance of patent scope: an empirical analysis. *The RAND Journal of Economics*, pages 319–333, 1994.

[52] Richard C Levin, Alvin K Klevorick, Richard R Nelson, Sidney G Winter, Richard Gilbert, and Zvi Griliches. Appropriating the returns from industrial research and development. *Brookings papers on economic activity*, pages 783–831, 1987.

[53] James Love. Evidence regarding research and development investments in innovative and non-innovative medicines. consumer project on technology, washington, dc. *Washington, DC*, 2003.

[54] Franco Malerba and Luigi Orsenigo. Knowledge, innovative activities and industrial evolution. *Industrial and corporate change*, 9(2):289–314, 2000.

[55] Edwin Mansfield. Academic research and industrial innovation: An update of empirical findings. *Research policy*, 26(7):773–776, 1998.

[56] Edwin Mansfield, Mark Schwartz, and Samuel Wagner. Imitation costs and patents: an empirical study. *The Economic Journal*, pages 907–918, 1981.

[57] Andreu Mas-Colell, Michael Dennis Whinston, Jerry R Green, et al. *Microeconomic theory*, volume 1. Oxford university press New York, 1995.

[58] MP Mathien. Parexel's pharmaceutical r&d statistical sourcebook 2002/2003. *Boston, MA, Parexel International Corporation*, 2002.

[59] Roberto Mazzoleni and Richard R Nelson. The benefits and costs of strong patent protection: a contribution to the current debate. *Research policy*, 27(3):273–284, 1998.

[60] Mariana Mazzucato and Giovanni Dosi. *Knowledge accumulation and industry evolution: The case of Pharma-Biotech*. Cambridge University Press, 2006.

[61] Jon F Merz and Mildred K Cho. What are gene patents and why are people worried about them? *Public Health Genomics*, 8(4):203–208, 2005.

[62] Petra Moser. How do patent laws influence innovation? evidence from nineteenth-century world's fairs. *American Economic Review*, 95(4):1214–1236, 2005.

[63] David Mowery and Nathan Rosenberg. The influence of market demand upon innovation: a critical review of some recent empirical studies. *Research policy*, 8(2):102–153, 1979.

[64] Kevin Murdock. Intrinsic motivation and optimal incentive contracts. *RAND Journal of Economics*, pages 650–671, 2002.

[65] Fiona Murray. Innovation as co-evolution of scientific and technological networks: exploring tissue engineering. *Research Policy*, 31(8):1389–1403, 2002.

[66] Summwe Myers and Donald G Marquis. *Successful industrial innovation*. Institute of Public Administration, 1969.

[67] Francis Narin, Kimberly S Hamilton, and Dominic Olivastro. The increasing linkage between us technology and public science. *Research Policy*, 26(3):317–330, 1997.

[68] Richard R Nelson and Bhaven N Sampat. Making sense of institutions as a factor shaping economic performance. *Revista de Economía Institucional*, 3(5):17–51, 2001.

[69] Richard R Nelson and Sidney G Winter. The schumpeterian tradeoff revisited. *The American Economic Review*, pages 114–132, 1982.

[70] Sean Nicholson. Financing research and development. *The Oxford Handbook of the Economics of the Biopharmaceutical Industry*, page 47, 2012.

[71] Alessandro Nuvolari. Collective invention during the british industrial revolution: the case of the cornish pumping engine. *Cambridge Journal of Economics*, pages 347–363, 2004.

[72] N Pattison and L Warren. drug industry profits: hefty pharmaceutical company margins dwarf other industries. washington (dc): Public citizen congress watch; june 2003, 2002.

[73] Keith Pavitt. The objectives of technology policy. *Science and public policy*, 14(4):182–188, 1987.

[74] Richard A Posner. An economic analysis of the use of citations in the law. *American Law and Economics Review*, 2(2):381–406, 2000.

[75] Derek J de Solla Price. The scientific foundations of science policy. 1965.

[76] Nathan Rosenberg. *Perspectives on technology*. CUP Archive, 1976.

[77] Jacob Schmookler. Invention and economic growth. 1966.

[78] Joseph Alois Schumpeter. *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle*, volume 55. Transaction publishers, 1934.

[79] Mariagrazia Squicciarini, Hélène Dernis, Chiara Criscuolo, et al. Measuring patent quality: Indicators of technological and economic value. Technical report, OECD Publishing, 2013.

[80] Scott Stern. Do scientists pay to be scientists? *Management science*, 50(6):835–853, 2004.

[81] Joseph E Stiglitz. Economic foundations of intellectual property rights. *Duke Law Journal*, pages 1693–1724, 2008.

[82] Joseph E Stiglitz. Intellectual property rights, the pool of knowledge, and innovation. Technical report, National Bureau of Economic Research, 2014.

[83] Catherine Tucker. Patent trolls and technology diffusion. 2013.

[84] Steven P Vallas, D Kleinman, and Dina Biscotti. Political structures and the making of us biotechnology. *F. Block and M. Keller (eds)*, 2011.

[85] Thorstein Veblen. Why is economics not an evolutionary science? *The Quarterly Journal of Economics*, 12(4):373–397, 1898.

[86] John P Walsh, Charlene Cho, and Wesley M Cohen. Patents, material transfers and access to research inputs in biomedical research. *Report to the National Academy of Sciences' Committee Intellectual Property Rights in Genomic and Protein-Related Inventions*, 2005.

[87] Sidney G Winter and Gabriel Szulanski. Replication of organizational routines. *The strategic management of intellectual capital and organizational knowledge*, pages 207–221, 2002.

[88] P Zuniga, D Guellec, H Dernis, M Khan, T Okazaki, and C Webb. Oecd patent statistics manual, 2009.